University of Zagreb

FACULTY OF PHARMACY AND BIOCHEMISTRY

Jelena Vlašić Tanasković

COMMUTABILITY EVALUATION OF CONTROL SAMPLES WITHIN EXTERNAL QUALITY ASSESSMENT PROGRAMS OF MEDICAL BIOCHEMICAL LABORATORIES

DOCTORAL DISSERTATION

Supervisors:
Wim Coucke, Ph.D.
Assoc. Prof. Jadranka Vuković Rodríguez

Zagreb, 2019.

Sveučilište u Zagrebu

FARMACEUTSKO-BIOKEMIJSKI FAKULTET

Jelena Vlašić Tanasković

PROSUDBA KOMUTABILNOSTI KONTROLNIH UZORAKA U PROGRAMIMA VANJSKE PROCJENE KVALITETE MEDICINSKO-BIOKEMIJSKIH LABORATORIJA

DOKTORSKI RAD

Mentori:
Dr.sc. Wim Coucke
Izv. prof. Jadranka Vuković Rodríguez

Zagreb, 2019.

The doctoral thesis was submitted to the Faculty Council of the Faculty of Pharmacy and Biochemistry, University of Zagreb in order to acquire a Ph.D. degree in the area of Biomedicine and Health, the field of Pharmacy, the branch of Medical Biochemistry

The work presented in this doctoral thesis was performed at the Croatian Centre for Quality Assessment in Laboratory Medicine, Croatian Society of Medical Biochemistry and Laboratory Medicine; Department of Laboratory Diagnostics, General Hospital Pula and Croatian Institute of Transfusion Medicine, Zagreb under supervision of Wim Coucke, Ph.D. and Assoc. Prof. Jadranka Vuković Rodríguez

ACKNOWLEDGMENTS

This thesis was a journey. An achievement so long expected that shaped me in so many ways. This feeling could not be achieved had it not been for so many of you who helped, advised, led and supported me throughout this process. You opened the world of science to me, and I'm so grateful I scratched the edges of it.

I'm immensely grateful to my supervisors, Mr. Wim Coucke, Ph.D. and Mrs. Jadranka Vuković Rodrìguez, Assoc. Prof, for their patience, constant support and numerous reviews of this thesis. Each one of you found the weak spots of this research and improved it to this final state.

My collaboration with Mr. Wim Coucke started several years ago, unexpectedly. He introduced me to the world of statistics, so vast and creative part of the science. He has shared his ideas and knowledge with me and many of the conclusions of this thesis are the result of our fruitful discussions. He helped me writing our first article and led me through all aspects of scientific progress that I'm now experiencing. I'm deeply indebted to his valuable and constant support.

My long-lasting friendship with Mrs. Jadranka Vuković Rodríguez has reached a new dimension. She found a way to guide and teach me, and her competence in presenting the results and ideas of this study were most valuable. She led me through my first proposal of this research and stayed persistent in her attempt to improve this thesis. Thank you Jadranka.

This thesis would not be possible without Mrs. Jasna Leniček Krleža, Ph.D., the Chair of the Croatian Centre for Quality Assessment in Laboratory Medicine. She reactivated already forgotten Ph.D. idea, believed in me, and she guided me throughout many steps in scientific progress.

My chief, Ms. Lorene Honović, Ph.D., the Head of Department of Laboratory Diagnostics in General Hospital Pula, was my constant support. She encouraged me in my pursuit in so many subtle ways and for such a long time. She opened many doors for me and her continuous motivation was many times priceless.

I have to thank Ms. Ana Hećimović, Ph.D., from the Croatian Institute of Transfusion Medicine, and her co-workers who helped me collect and prepare fresh serum samples for this research. I'm also thankful to Mrs. Snježana Hrabrić Vlah, Mr. Goran Ferenčak, Ph.D. and all laboratory personnel who helped to analyse so many samples.

My special thanks go to my mom and my whole family; they have always been there for me and especially throughout these busy times.

My love goes to my children, Ana and Vito, and my husband Petar. I dedicate this work to them. They are my strength and my joy. In time, my children will understand the need for continuous education and personal development that I was aiming at. My husband does, as he has proven so many times.

SUMMARY

Introduction and aim: External quality assessment (EQA) is an integral part of quality management systems in medical biochemical laboratories enabling monitoring of individual results as well as harmonisation and standardisation of measurement procedures (MPs) used in the clinical setting. Commutability of control samples is a major prerequisite for assessing laboratory and MP performance according to the unique target value. Commutable control samples show the same properties in different MPs as well as patient samples. Commutability is usually evaluated using regression analysis and statistically determined criteria of acceptance without taking into consideration analytical performance specifications for the analyte. The aim of this research is to propose a new model for the evaluation of commutability criteria using analytical performance specifications for each analyte within the EQA program for medical biochemical laboratories.

Materials and methods: Lyophilised control samples were distributed together with native and spiked serum samples to all participants of Croatian EQA (CROQALM). The participants analysed both samples using routine MPs. Commutability of control samples was evaluated using the results of two kinds of samples and newly proposed false flagging method. The results for commutability were compared to statistically determined commutability criteria obtained by recommended regression analysis for commutability evaluation of EQA control samples. Three lyophilised EQA control samples were evaluated for commutability for 22 biochemistry analytes and related MPs used in medical biochemical laboratories.

Results: The controls were found commutable for 13 analytes: AMY, AST, CK, glucose, iron, LDH, phosphate, potassium, sodium, proteins, triglycerides, urate and urea. High noncommutability of control materials was found for chloride in all three control samples and HDL-cholesterol, AP, creatinine and calcium in two out of three control samples. Unequal criteria in statistically defined commutability limits resulted in commutability conclusions that are dependent on measurement results of patient serum samples by evaluated MPs.

Conclusions: The false flagging method, proposed in this thesis, can be used for evaluating commutability of control samples within the EQA program of medical biochemical laboratories. The commutability limits are equally designed for all MP combinations and connected to established analytical performance specifications of the analytes.

Keywords: commutability, external quality assessment, false flagging method

SAŽETAK

Uvod i cilj: Vanjska procjena kvalitete sastavni je dio sustava za upravljanje kvalitetom medicinsko-biokemijskih laboratorija. Osim prosudbe mjernih rezultata, vanjska procjena kvalitete ima za svrhu praćenje globalnih ciljeva harmonizacije i standardizacije mjernih postupaka koji se koriste u laboratorijima. Cilj takvog praćenja je osiguranje mjeriteljske sljedivosti rezultata analiza te mogućnost da se koriste jedinstveni referenti materijali i slijede istovrsne kliničke smjernice. Komutabilnost kontrolnih uzoraka nužan je preduvjet za valjanu prosudbu kvalitete prema jedinstvenoj ciljnoj vrijednosti, a definirana je kao bliskost numeričkog odnosa između rezultata različitih mjernih postupaka za referentni materijal kao i za reprezentativne uzorke pacijenata, ovisno o namjeni referentnog materijala. Premda proizvođači i programi vanjske kontrole kvalitete nastoje osigurati komutabilne uzorke za prosudbu laboratorija, komutabilnost je vrlo često ugrožena zbog nastojanja da se osiguraju dovoljne količine kontrolnog uzorka stabilnog kroz duže razdoblje i koji sadrži različite koncentracijske raspone ispitivanih analita. Metode koje se najčešće koriste za ispitivanje komutabilnosti temelje se na regresijskoj analizi i na usporedbi kontrolnih uzoraka s uzorcima pacijenata uz interval pouzdanosti od 95% oko linije regresije kao kriterija prihvata. Statistički kriteriji za prosudbu komutabilnosti omogućavaju objektivnu, brojčanu prosudbu rezultata mjerenja, no kriteriji prihvata u velikoj mjeri ovise o stupnju usporedivosti dvaju mjernih postupaka na uzorcima pacijenata. Do sada predloženi statistički kriteriji ne uzimaju u obzir svrhu korištenja ispitivanih kontrolnih uzoraka, te ciljeve analitičke ili kliničke kvalitete za pojedini analit. Stoga je cilj ovog doktorskog rada postavljanje i validacija nove metode za prosudbu komutabilnosti kontrolnih uzoraka kojom se komutabilnost kontrolnih uzoraka prosuđuje ovisno o postavljenim analitičkim ciljevima kvalitete i njihovoj konačnoj namjeni procjene točnosti rezultata mjerenja i standardizacije/harmonizacije mjernih postupaka.

Materijali i metode: U ovom istraživanju korišteni su svježi serumi dobrovoljnih davatelja krvi; svježi serumi dobrovoljnih davatelja krvi s dodatkom glukoze, ureje, natrija, kalija, klorida i bilirubina, ostatni uzorci seruma pacijenata koji se prikupljaju nakon rutinske laboratorijske obrade, te tri liofilizirana komercijalna kontrolna uzorka (C1/2016, C2/2016 i C3/2016) različitih proizvođača koji se koriste u vanjskoj procjeni kvalitete medicinsko-biokemijskih laboratorija u Hrvatskoj. Rezultati mjerenja 12 ispitivanih analita (glukoze, ukupnog kolesterola, triglicerida, HDL-kolesterola, ureje, kreatinina, natrija, kalija, klorida, AST, ALT i GGT) u liofiliziranim kontrolnim uzorcima uspoređivani su s rezultatima mjerenja istih analita u ostatnim serumima pacijenata upotrebom pet rutinskih mjernih postupaka. Prvi

korak u prosudbi komutabilnosti kontrolnih uzoraka bila je regresijska analiza. U okviru vanjske procjene kvalitete medicinsko-biokemijskih laboratorija, kontrolni uzorci i serumi dobrovoljnih davatelja krvi analizirani su u 180-184 medicinsko-biokemijska laboratorija tijekom 2016. godine, korištenjem standardnih mjernih postupaka, u tri ciklusa vanjske procjene kvalitete CROQALM. Analiza uzoraka obuhvatila je mjerenje svih biokemijskih pretraga obuhvaćenih ovim programom koje ulaze u opseg rada danog laboratorija. Dobiveni rezultati grupirani su prema mjernim uređajima i metodama u 143 mjerna postupka koji su korišteni za mjerenje 22 analita: glukoza, ureja, kreatinin, bilirubin, urati, natrij, kalij, kloridi, kalcij, ukupni kolesterol, trigliceridi, HDL-kolesterol, AST, ALT, AP, GGT, CK, LDH, amilaze, željezo i ukupni proteini. Procjena statistički značajnih razlika između rezultata mjerenja kontrolnih uzoraka i uzoraka seruma provedena je analizom varijance (ANOVA). Kako bi se omogućila analiza velikog broja uzoraka i MP, predložena je i razvijena nova metoda, tzv. metoda lažnog odstupanja (engl. *false flagging method*), kojom se prosuđuje komutabilnost kontrolnih uzoraka. Metoda se temelji na određivanju najvećeg dopuštenog udjela odstupanja u prolaznosti laboratorija na kontrolnim uzorcima u usporedbi s udjelom prolaznosti na uzorcima seruma. Rezultati prolaznosti laboratorija prema zadanim ciljevima kvalitete za svaki analit uspoređivani su za svaku vrstu uzorka u pojedinom ciklusu distribucije (kontrolni uzorak i serum).

Rezultati: Korištenjem regresijske analize, sva tri kontrolna uzorka pokazala su komutabilnost za ispitivane parove mjernih postupaka koji se koriste za mjerenje kalija, natrija, GGT, AST i triglicerida. Nekomutabilnost je dokazana za kolesterol, HDL-kolesterol i glukozu u sva tri kontrolna uzorka te kloride u kontrolama normalnog i kreatinina visokog koncentracijskog raspona ispitivanog analita. Nekomutabilnost kontrolnog uzorka C3/2016 dokazana je za većinu usporedbi između parova mjernih postupaka za ALT. Kako bi se utvrdila statistički značajna razlika između mjerenja dobivenih na kontrolnim uzorcima i uzorcima seruma u istoj seriji na uređaju, u okviru vanjske procjene kvalitete medicinsko-biokemijskih laboratorija, uspoređivani su rezultati mjerenja obje vrste uzoraka analizom varijance. Dobiveni rezultati upućuju na postojanje statistički značajnih odstupanja između kontrolnih uzoraka i uzoraka seruma za 22 – 36,1% parova mjernih postupaka ovisno o vrsti kontrole. Sve tri kontrole pokazuju komutabilnost za kalcij, CK, proteine i ureju, a nekomutabilnost za većinu kombinacija mjernih postupaka za mjerenje klorida i HDL-kolesterola. Primjenom nove predložene metode za prosudbu komutabilnosti kontrolnih uzoraka, kontrolni uzorci prosuđivani su prema postavljenim analitičkim ciljevima kvalitete za svaki analit. Metodom lažnog odstupanja ispitana je komutabilnost kontrolnih uzoraka za 22 analita i 331-426 parova

mjernih postupka koji se koriste u rutinskom radu laboratorija. Sva tri kontrolna uzorka pokazuju komutabilnost za većinu kombinacija mjernih postupaka za mjerenje amilaze, AST, CK, glukoze, željeza, LDH, fosfata, kalija, natrija, proteina, triglicerida, urata i ureje. Nekomutabilnost sva tri kontrolna uzorka dokazana je za kloride, te HDL-kolesterol, AP, kreatinin i kalcij u dvije kontrole. Sveukupno, kontrolni uzorci Seronorm Human (C1/2016 i C2/2016) proizvođača SERO pokazuju veći ukupni postotak komutabilnosti za ispitivane analite i mjerne postupke (83,1% i 87,6%) od kontrolnog uzorka C3/2016 proizvođača Fortress Diagnostics (76,1%).

Zaključci: Postupak regresijske analize za procjenu komutabilnosti kontrolnih uzoraka koji se koriste u programima vanjske procjene kvalitete, organizacijski je i financijski zahtjevan zbog velikog broja analita koje treba ispitati za sve mjerne postupke koji se rutinski provode u medicinsko-biokemijskim laboratorijima. Osim toga, kriteriji prosudbe komutabilnosti koji se koriste u regresijskoj analizi ovise o statističkim značajkama dobivenih rezultata i različiti su za svaku ispitivanu kombinaciju mjernih postupaka. Primjenom metode lažnog odstupanja istovremenom analizom kontrolnog uzorka i uzorka svježeg seruma na velikom broju mjernih postupaka, moguća je prosudba komutabilnosti kontrolnih uzoraka u okviru sheme vanjske procjene kvalitete. Utvrđivanjem najvećeg dopuštenog udjela lažnog odstupanja rezultata mjerenja kontrolnog uzorka od rezultata mjerenja na uzorku seruma, komutabilnost kontrolnih uzoraka prosuđuje se na temelju razlike udjela prolaznosti laboratorija na dvije vrste uzoraka. Ukoliko je udio prolaznosti laboratorija značajno različit na kontrolnim uzorcima u usporedbi s uzorcima seruma, potvrđuje se različito ponašanje kontrolnih uzoraka od uzorka seruma na istim mjernim postupcima, odnosno nekomutabilnost kontrolnih uzoraka. Ovim postupkom su kriteriji prosudbe jednoznačni za sve parove mjernih postupaka, omogućavajući prosudbu kliničke i/ili analitičke jednakovrijednosti kontrolnih uzoraka prema dijagnostičkim značajkama samog analita. Metoda lažnog odstupanja predložena u ovom radu predstavlja novi pristup u prosudbi komutabilnosti i može se primijeniti istovremeno za veliki broj analita i mjernih postupaka u okviru vanjske procjene kvalitete medicinsko-biokemijskih laboratorija.

Ključne riječi: komutabilnost, vanjska procjena kvalitete, metoda lažnog odstupanja

TABLE OF CONTENTS

1. INTRODUCTION

Laboratory diagnostics plays an important role in overall patient management and is often included in diagnosis, follow-up and treatment of various diseases (1). The number and the variety of laboratory tests performed in medical biochemical laboratories increases over time and the results obtained in the laboratory regularly serve as a basis for clinical decision making. In order to meet high standards regarding patient safety and medical care, quality management of the total testing process (TTP) became an indispensable part of laboratory medicine (2,3).

The purpose of laboratory quality management is validation, implementation and monitoring of all pre-analytical, analytical and post-analytical processes in the laboratory, thus identifying key quality indicators to be evaluated and very often improved over time. Assessment of laboratory performance and quality of total TTP is usually validated through guidelines and regulations provided by national and international regulatory bodies, such as Clinical Laboratory Improvement Amendments (CLIA), Guideline of the German Medical Association on Quality Assurance in Medical Laboratory Examinations (RiliBÄK), Croatian Chamber of Medical Biochemists (CCMB) and ISO 15189:2012 (4–7).

Quality assessment of the analytical part of TTP relies mainly on data from internal quality controls (IQC) and external quality assessment (EQA) programs. In addition to the validation and/or verification of the measurement procedures (MPs) used in medical biochemical laboratories (MBL) and regular performing of IQC, participation is EQA programs is nowadays an “integrated professional activity of medical laboratories”, providing quality assessment and bases for improving activities ensuring high-quality standards in medical care for the patients (8,9).

1.1 External quality assessment

1.1.1 General aspects

External quality assessment (EQA) was recognised more than half a century ago as a tool to recognise methods with poor performance in an interlaboratory comparison survey described by Belk and Sunderman in 1947 (10). Initially conducted only for several analytes, the EQA evolved in forthcoming years in a number of surveys and scope and was recognised by professionals as an essential component of quality management. The term *external quality assessment* is used to describe the method or process that allows comparison of laboratory's testing to that of a source outside the laboratory – peer group of laboratories or reference laboratory (11). The term is very often used interchangeably with *proficiency testing* (PT), however, EQA usually implies broader spectrum of quality assessment, including educational, supportive and structured approach towards improvement in laboratory performance (12,13). Although traditionally addressing analytical quality, EQA can be applied to other aspects of total testing, both pre-analytical and post-analytical processes (14–16). Participation in an EQA program provides objective assessment and information on performance and quality of results delivered to patients and physicians. It helps to monitor individual laboratory performance over time, identifying problems in analytical and extra-analytical processes, gives information on the suitability of diagnostic systems, the accountability and competence of the laboratory staff and indicates areas that need improvement (17,18). In terms of analytical performance, it provides information on the reliability of applied methods and equipment as well as the validity of uncertainty claims. Over time, participation in EQA program can lead to an improvement in the quality of laboratory performance, assuming monitoring and root causes of any discrepancy in EQA result are properly addressed and actions toward improvement taken (19,20). The information from EQA reports can be used to reduce the bias of the methods, confirm the quality of results and increase the confidence in laboratory performance (21). It also serves as a compliance proof for a laboratory's ability to meet aimed quality standards, often the subject of close inspection from various regulatory and accreditation bodies.

1.1.2 Harmonisation and standardisation in laboratory medicine

In addition to individual laboratory evaluation, EQA has a central role in monitoring and promoting global initiatives towards standardisation and harmonisation of laboratory results (21–23). Comparable, or harmonised, test results across different measurement systems, laboratories, time and locations becomes an important activity of scientific and professional community (24,25). The underlying reason for all harmonisation efforts is an overall benefit for patients who are often diagnosed and treated across different medical facilities, even health care systems, where the results from the laboratories are shared between those. In such perspective, test results must be harmonised or equivalent between laboratories allowing the use of same evidence-based clinical guidelines, reference intervals and decision levels in interpreting results. For example, using internationally accepted guidelines such as *Kidney Disease Improvement Global Outcomes* (KDIGO) guidelines for the diagnosis and management of chronic kidney disease is valid only if the results for creatinine from the patient laboratory are comparable to the results of laboratories used in the clinical studies (14,26). In addition, harmonisation of test results also raises the level of confidence in laboratory diagnostics and diminishes confusion of both doctors and patients. As Plebani (27) observed in terms of present differences in measurement and cut-off limits for troponins in acute myocardial infarction, it should be possible to diagnose acute myocardial infarction irrespective of the choice of analyte (cardiac troponin I or cardiac troponin T) and analyser.

A very important aspect of harmonisation in consolidation and networking is the benefit of sharing patient results by a wide range of users across different levels of the healthcare system, often as a part of patient's electronic record (28,29). The need for harmonisation goes even beyond methods and analytes, and includes all parts of TTP (27,30).

Harmonisation in measurements from different analytical systems is commonly achieved through standardisation and traceability of all procedures to a higher-order reference system (31–33). Reference materials (RMs) are defined in ISO documents as *materials, sufficiently homogenous and stable with respect to one or more specified properties, which have been established to be fit for their intended use in a measurement process* (34). Although closely linked and often used interchangeably, harmonisation and standardisation refer to two distinct concepts in metrology principles. Standardisation implies traceability of results reported in SI units (*Système International Units*, SI) to higher-order RMs and/or methods, whereas harmonisation means consistency, or comparability of measurement results (24,27).

Comparability in measurement results can be achieved by standardisation for defined chemical entities, traceable to SI units. For heterogeneous, complex analytes not directly traceable to SI units, where neither higher order primary RM and/or method exist, harmonisation can be achieved either by consensus traceability to some reference or comparison between methods following mathematical corrections (24,35,36). For example, pursuing harmonisation trough standardisation is possible for rather "simple" analytes such as glucose, electrolytes or cholesterol, but challenging for complex heterogeneous analytes such as troponins, tumour markers and many others. It however has to be noticed, that although in a minority, those "simple" analytes represent the most commonly requested tests in medical biochemical laboratories (22).

A very important step in implementing standardisation as a principal method in achieving harmonisation of measurement results is enforcement of the *In Vitro Diagnostic Directive* (IVDD) (33) from 1998 which requires manufacturers of diagnostic devices with CE (*Conformité Européene*) mark to provide traceability for assays and calibrators. Basic concepts and procedures are further defined and specified in ISO 17511:2003 (37). The calibration transfer protocol, as described in ISO 17511, is presented in Figure 1.

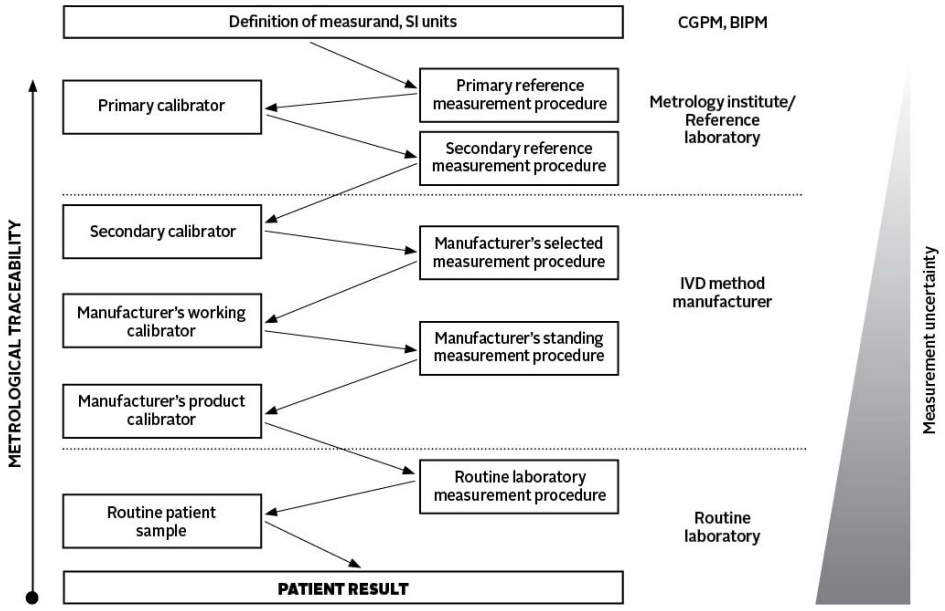


Figure 1. Calibration transfer protocols for cases with primary reference MPs and primary calibrators giving metrological traceability to SI. Abbreviations: ARML, Accredited reference measurement laboratory; BIPM, Bureau International des Poids et Mesures; CGMIP, Conférence Générale des Poids et Mesures; ML, Manufacturer's laboratory; NMI, National Metrology Institute; uc (γ), uncertainty. (Modified according to reference 37.)

It can be seen that primary RM can be prepared from chemically pure substance using primary reference procedure such as gravimetry. Such material further serves as a calibrator for secondary reference MP, which, in turn, is used to assign a true value to secondary RM used by manufacturers. It should be noted here that the secondary reference procedure is insensitive to matrix differences between its calibrator and secondary reference calibrator to be used by manufacturers of instruments and/or reagents. On this level, after being calibrated by secondary reference calibrator, manufacturers usually assign a value to their working calibrator or master calibrator. It further serves as a calibrator for end-users MPs in MBLs. Each of these steps in hierarchically organised traceability chain has its measurement uncertainty, resulting in a combined overall uncertainty of the end-user's calibrators and patient results. Measurements of cholesterol and HbA_{1c} are examples of successful standardisation processes with consequential clinical impact (38). However, even standardisation and traceability to higher-order reference systems must be monitored and acceptable measurements uncertainties fit for clinical use have to be defined (39,40). Otherwise, the theoretical benefit of the whole traceability process might be absent, resulting in the poor harmonisation of results due to different types of metrological chains used by manufacturers with large "grey zones" regarding acceptable measurement uncertainties across the traceability protocol (41,42). Achieving harmonisation is a global activity that needs active involvement from all stakeholders, i.e. metrologists, international standards organisations, IVD method manufacturers, regulation/accreditation bodies, EQA providers and medical biochemical laboratories (43). In those terms, EQA is recognised as an important and powerful tool in monitoring and supporting harmonisation and standardisation in laboratory medicine (14,22,31). In order to support worldwide comparability and harmonisation, the *Joint Committee for Traceability in Laboratory Medicine* (JCTLM) was formed as an international committee in 2002 by *Bureau International des Poids et Mesures* (BIPM), *International Federation of Clinical Chemistry and Laboratory Medicine* (IFCC) and *International Laboratory Accreditation Cooperation* (ILAC), bringing together governmental organisations, clinical laboratory professionals and the IVD industry (44). JCTLM recognised three *pillars* in standardisation and metrological traceability: higher-order RMs, higher-order reference methods and accredited reference laboratory services. In addition to forming the web-based database of higher-order materials, methods and reference laboratory services, JCTLM promotes and actively encourages all traceability concepts in agreement with internationally accepted standards, recognises and objectively evaluates new materials and methods and provides educational material for all stakeholders involved (45,46). In addition to the three pillars identified by JCTLM, laboratory

professionals identified three more: universal reference intervals and medical decision levels, EQA programs using commutable samples with reference method target values, and limits for uncertainty and error of measurement fit for clinical use (23,39,40,47,48). EQA is thus recognised as an indispensable tool in verifying performance and the quality standard achieved in a participating laboratory, but also in monitoring and promoting metrological traceability, standardisation and harmonisation of laboratory results.

1.1.3 Principal characteristics of the EQA program and survey design

An EQA program can be organised in a national, international or regional level depending on the participating laboratories and the demands of various governmental, healthcare or professional agencies. Furthermore, the various EQA programs differ significantly in terms of the organisation; the scope of the program (analytical, pre-analytical and post-analytical phase of laboratory work), variety of tests offered, number of EQA surveys per year, the obligation of participation in the program, evaluation particularities, etc. In order to meet the intended use of the EQA in quality improvement and education, EQA providers share the knowledge and cooperate to constantly improve their service to participants and are often governed, even evaluated according to various international guidelines and standards (11,17,49,50).

The usual EQA survey is conducted by sending a set of samples with an unknown concentration of one or many analytes to participating laboratories, together with instructions on proper handling, preparing and analysing the samples (Figure 2). According to given instructions, participating laboratories perform the analysis of received samples as if they were patient samples and send the results back to scheme organiser. The scheme organiser collects and evaluates data sent from participants to create EQA reports, important feedback tool for laboratories. The reports should be understandable and comprehensive, containing information on assigned values and analytical performance specifications for specific measurand, supported by the graphical presentation of laboratory's results compared to the results of other laboratories (51). The reports usually contain the evaluation analysis on laboratory performance, as well as the method and/or instrument performance based on the results from many laboratories. Every laboratory is expected and encouraged to follow up any inconsistency or unacceptable EQA result, find a root cause to inconsistency or unacceptable result, take corrective actions and document changes (13,52). Many schemes provide a graphical

presentation of laboratory performance over time, thus enabling laboratories to follow up the quality of their laboratory procedures and evaluate new trends in terms of deterioration or improvement observed.

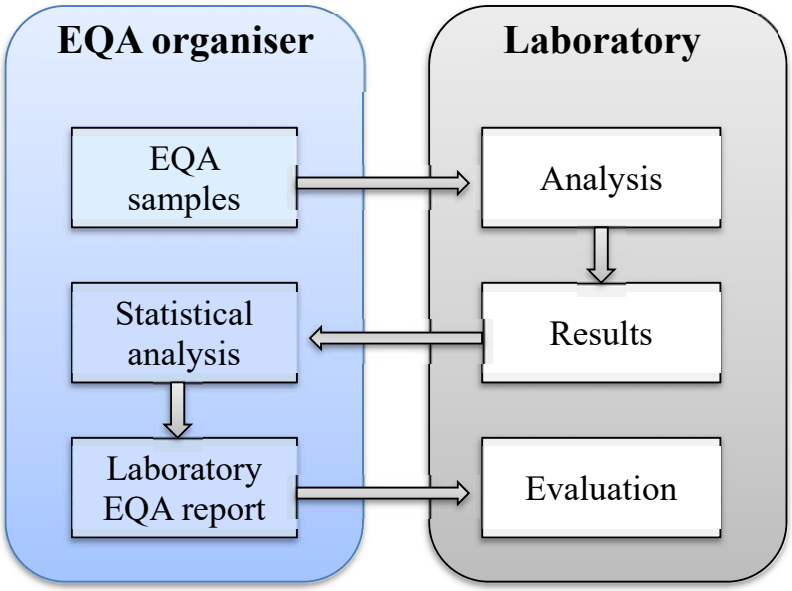


Figure 2. The flowchart of an EQA survey.

1.1.4 Interpretation of results within the EQA program: analytical performance specifications and target values

Analytical performance specifications. The key elements in results evaluation within the EQA program are target values and acceptance limits around those values, or analytical performance specifications for the measurand. Analytical performance specifications should be defined prior to result analysis and criteria or rationale for their setting must be clear to participants. This way the laboratories can have confidence in the scheme and are informed on the quality level needed or achieved in EQA (51,53,54). Analytical performance specifications differ largely in various EQA schemes and it is quite possible that individual result or quality level achieved in the laboratory might be considered differently by these schemes in terms of

fulfilling appropriate quality standards (14,55). The terminology used to describe allowed deviations from the assigned values is also different throughout literature and EQA programs, referred to as Analytical Performance Specifications, Allowable Limits of Performance, Acceptability Limits, and Quality Goals. The term Analytical Performance Specifications (APS) is preferred and adopted by *European Federation of Clinical Chemistry and Laboratory Medicine* (EFLM), *Task and Finish Group on Performance Specifications for EQAS* (TFG-APSEQA) to be in the line of the terminology used in Milan strategic conference on analytical performance goals in 2014 (56). The Milan conference was a follow-up conference held by EFLM to revise the original hierarchy of APS established in Stockholm (57). The structured approach criteria in setting APS in laboratory medicine originally proposed in so-called Stockholm criteria is somewhat shortened and simplified in Milan, and three models for establishing APS were suggested (Table 1).

Table 1. Recommended models in setting analytical performance specifications

Model	Bases on which different models for APS are set
1	Effects of test performance on clinical outcome Direct outcome studies – investigating the impact of the performance of the test on clinical outcome Indirect outcome studies – investigating the impact of the performance of the test on clinical classification or decision
2	Components of biological variation of the measurand
3	State-of-the-art of the measurement – the highest level of analytical performance technically achievable

Hierarchically organised, the criteria are based on the clinical outcome, components of biological variation and *state-of-the-art*. The preferred model for setting APS is a model based on the expected effect on clinical outcome, coming from direct or indirect clinical studies. Although this model is set on the top of the hierarchy, clear evidence by randomised control trials on the effect of established APS on clinical outcome is still lacking (58). However, outcome-related studies reflect the clinical needs of patients and should be encouraged. The model based on components of biological variation is the most widely used model in establishing APS. The database of desirable, minimum and maximum quality specifications is hosted at http://www.westgard.com and future updates are set to be handled by EFLM (59,60). The third model, the model based on the *state-of-the-art,* is the highest level that can be

achieved using current technology. Although the models are distinct in their basic principle, they can be used simultaneously, for example, a *state-of-the-art* model can be chosen to set desirable, optimal or minimal criteria from the biological variation of specific measurands (61). Criteria for assigning measurands to different models largely depend on the role of the measurand in a clinical setting (diagnosis, monitoring) and the ability of IVD industry and laboratories to meet different levels of quality (62). Furthermore, the level of quality depends on the expected response by participants to failure, and can be set by EQA scheme as passable or satisfactory (favoured approach for regulatory requirements), favourable (where further improvement is not needed) and aspirational (aiming at improving quality or performance) (53).

Target values. The target value is another key element when assessing individual performance through the EQA program since every result is compared to that particular value. In order to evaluate laboratory performance, results are usually presented as the difference between laboratory result and the target value (D-score), expressed as a percentage, thus allowing comparison with established APS (17). Following this criterion, and regardless of the choice or rationale used for setting APS, a laboratory result is 'flagged' if the relative deviation from target value exceeds allowed APS.

Z-scores are also commonly used through EQA for evaluation of the individual result. They are the difference between the laboratory result and target value corrected for variability (51). The Z-score is sometimes referred as statistically-based acceptance criterion, where scores with an absolute value below 2 are considered as acceptable, between 2 – 3 questionable ("warning signal") and Z-scores greater than 3 are considered unacceptable (13,17). Very often, the performance is evaluated by a combination of performance scores, supported by a graphical presentation of results and interpretative comments from the EQA provider to sustain the educational role of EQA.

The example of one EQA evaluation report for individual laboratory and analyte is given in Figure 3. It shows the participant's results of the iron analysis in two EQA samples. The top two graphs present the histograms of all data submitted with the laboratory's method group separated from all groups with a different colour. The result reported by the laboratory is presented with a red dot on the histogram and numerically underneath the graph, together with the percentage deviation from the target value (X_T). The statistical analysis of the laboratory's method group and all results submitted are shown below the histograms. The graphs on the bottom present current and the previous results with the green-shaded area of acceptance limits in percentage (bottom left) and absolute (bottom right) deviations from the target value. These

graphs show the laboratory performance over a longer period of time and can be used to detect any long-term bias.

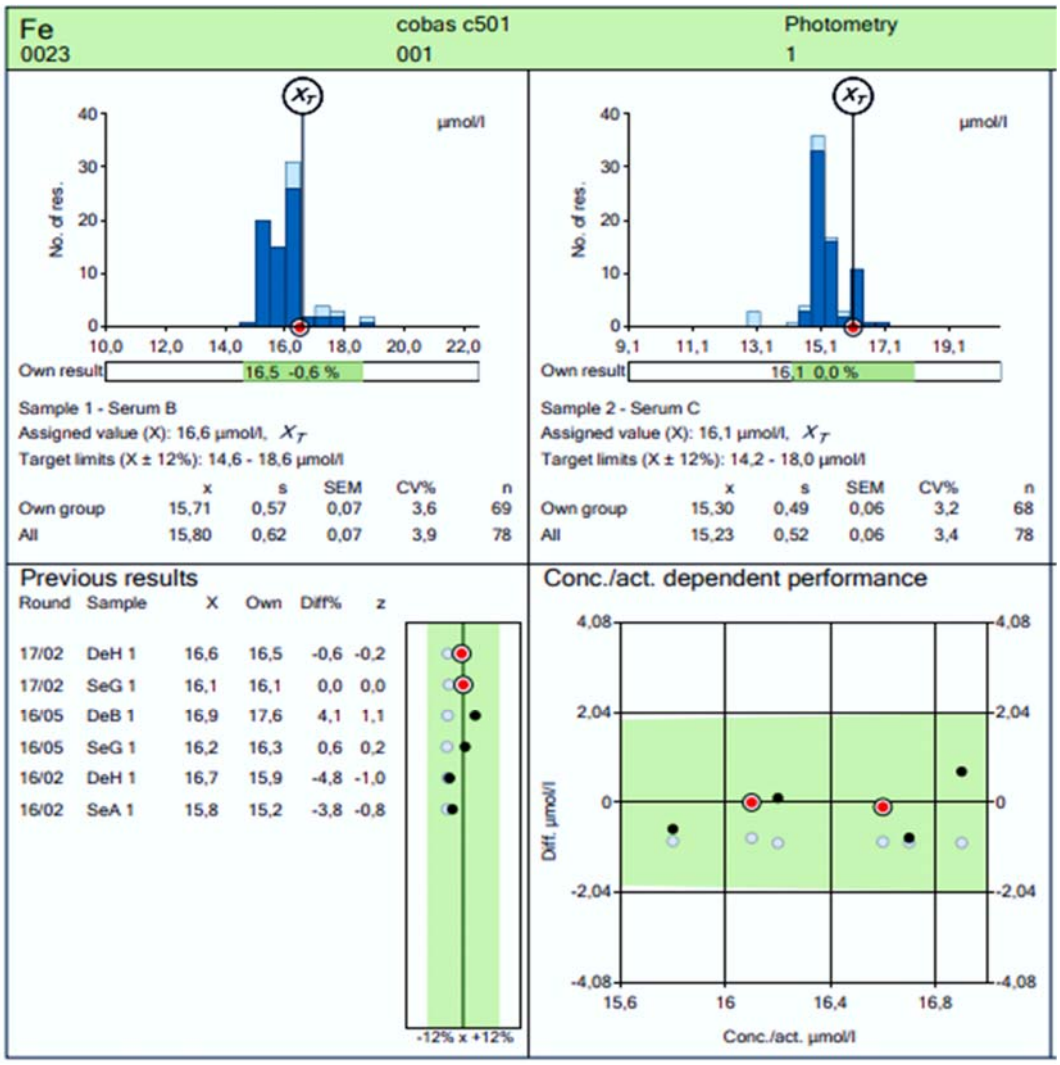


Figure 3. Laboratory EQA report for iron analysis in two control samples. X_T - assigned target value; x – consensus mean value; s – standard deviation; SEM – standard error of the mean; CV% - coefficient of variation, n – number of reported results, Diff% - percentage deviation from assigned target value, Diff. mmol/l – absolute deviation from assigned target value. Dark blue bars in the histogram represent the results from the laboratory's (own) peer group and light-blue rectangles represent all results. Green-shaded areas in the bottom two graphs represent the acceptance limits in percentages deviations and z-scores (bottom left) and absolute deviations from target value (bottom right). The results from the current EQA survey are presented with red dots and the results from the previous surveys with black dots. The grey dots indicate the laboratory's peer group consensus mean.

The choice of the target value is very important when assessing the distance of received results from the target value and in calculating various performance scores, like D-score or Z-score. EQA organizers have used two types of target values: consensus target values and assigned target values. The essential difference is that consensus values are derived from reported results and are determined using statistical calculations for estimation of a central value, whereas assigned target values are known to EQA organizers beforehand and are not dependent on participants' results. Consensus values can be calculated from all participants in a homogenous population, assuming correct use of statistical techniques and methods to solve major issues that might jeopardize correct statistical evaluation such as the exclusion of outliers, bimodality and skewness (51,63). The commonly used consensus target values are robust estimators of a central value, such as median and "all method trimmed mean", mostly depending on the particular choice of the EQA organizer (50,64). The consensus value can be also derived from results obtained from "best performing laboratories" or few laboratories chosen by EQA organizer. The assigned target value is ideally obtained by analysing the EQA samples in a reference laboratory using the reference method. The list of such laboratories and services is provided by JCTLM in order to support traceability and standardization of MPs to higher-order RMs. The reference value in some EQA programs is assured using a transfer protocol by which selected laboratories are measuring both certified RM and EQA sample, and the target value is determined after correction of observed bias from RM (65). EQA programs with target values assigned by reference methods and materials allow accuracy-based evaluation of both laboratories and MPs on the market. In order to fit for that purpose, commutability of EQA samples must be validated to ensure that the difference from the assigned target value is caused by calibration bias rather than matrix-related bias (52,66). When commutability is not assessed or reference MPs are not available, the choice of the target values is restricted to consensus target values in peer-groups which are expected to have the same result for particular EQA sample (67). Hence, besides the availability of applicable references, it is the quality and characteristics of EQA samples that mainly determine the choice of target values and evaluation capabilities of EQA (23,52,68)

1.1.5 The characteristics of EQA samples

EQA samples can be prepared by EQA organizers or acquired from an external source, usually commercial suppliers of control materials. Regardless of the source of the samples, they must be suitable for clinical use and cover the analytical range of interest, usually in the low, "normal", and high levels compared to the reference interval of an analyte. Furthermore, every laboratory should get substantially equal sample material for analysis; so, homogeneity and stability must be assured for the time samples are transported and analysed by participants. Since the samples are only one part of an EQA program, the expenses for their preparation or purchase have to be reasonable and affordable by participating laboratories. Above all, considering the fact that EQA samples have to be used as routine samples, they should behave in the same manner as patient samples in laboratory MP, i.e., they should be commutable. Fulfilling all of those requirements is very demanding in practice, and some compromises are usually necessary for the preparation of EQA samples. The most important characteristic of EQA samples is commutability with patient samples, very often being contrary, or even antagonistic to other criteria. In other words, in the pursuit of samples with acceptable stability, concentration, price and other requirements for ideal EQA sample, commutability of control samples is often compromised (52,69). Every intervention in authentic human samples like spiking (supplementation with analytes), pooling, freeze-thaw cycles, lyophilisation, filtration, etc. can lead to noncommutability with authentic patient samples. Various manufacturing procedures cause matrix modifications, which in turn can lead to alternations of physical and chemical properties of one or more components or introduce non-native molecules. The matrix here is defined as the total of all components of the material except the analyte itself (37). For example, lyophilisation irreversibly denaturates lipoproteins, causing modifications in viscosity, turbidity, pH and surface tension (70,71). The difference from patient samples is sometimes the result of changes in analyte rather than the matrix, like the addition of enzymes from the non-human origin which sometimes have different properties than human enzymes like optimal substrate and pH, the effect of inhibitors, etc. (70,72). Even minor interventions in serum preparation like sterile filtration, storage before aliquoting and freezing may disturb the equilibrium between protein-bound and free thyroid hormone and endanger commutability (73).

It has been commonly agreed that minimally altered or processed off-the-clot serum samples are likely to be commutable with patient samples, and the validity of such assumption is mostly

based on the stringency of their preparation (52,66,69,74). Single-donation serum or pooled serum samples may be used, due to the fact that high volumes are usually needed and the possibility that interferents present in single-donation serum may influence commutability (69). On the other hand, pooling the samples may introduce further interactions and complex formation between different components in serum and thus compromise the original characteristics of native serum samples. It has been hypothesized and further reported that supplementation with purified simple analytes doesn't influence the commutability of EQA material (70,74). This assumption has to be taken with caution, since more complex analytes may not behave in the same manner or even be obtained in highly purified forms. Every artificial procedure and intervention applied to native clinical specimens may introduce noncommutability of samples, causing changes in reactivity through matrix-sensitive procedures, such that measurement characteristics are no longer representative of patient samples. It is thus important to verify the commutability of EQA samples used to simulate closely relevant properties of patient samples intended to be measured.
Thus, commutability with clinical patient samples is one of the most important concepts affecting the design and interpretation of EQA programs.

1.2 Commutability

1.2.1 Definitions and description

Commutability is the property of RMs indicating the same inter-assay relationship of those materials and authentic patient samples. RMs hereby refer to all materials used to calibrate a MP or to assess the trueness of measurement results, including calibrators used in medical biochemical laboratories, trueness controls and certified RMs (75). To be able to serve as calibrator or trueness control in certain steps of metrological traceability chain, commutability of RM has to be assessed, and fitness for the intended use established (76). The term commutability was initially used to describe the ability of control materials to show the same characteristics as patient samples in different MPs for enzymes, and it was later expanded to

other analytes (77,78). Several definitions of commutability are used throughout scientific literature and standard documents. ISO documents define commutability as *the equivalence of mathematical relationship between the results of different MPs for a RM and for the representative samples from healthy and diseased individuals* (37). The International Vocabulary of Metrology (VIM) states that commutability is a *property of RM, demonstrated by the closeness of agreement between the relation among the measurement results for a stated quantity in this material, obtained according to two given MPs, and the relation obtained among the measurement results for other specified materials*, further noted as routine samples (79). Basic principles in both definitions are similar, and, translated in common language; the commutability describes the same behaviour of RM as native patient samples in different MPs. Although the property of a RM, commutability is in fact attributed to analyte-material-method interaction, and a specific material can be found commutable for some analytes and methods, and noncommutable for others. For example, RM ERM-DA470k/IFCC used as the common calibrator for serum proteins was found commutable for all proteins except C-reactive protein (CRP) and ceruloplasmin (80,81). Commutability of a RM goes even beyond analytes and methods and includes even specific reagent lots interactions (82). It is thus common to evaluate commutability of RM for specified MP, which includes method specifications, instrument and reagents in use. Noncommutability is sometimes referred to as *matrix-effect* or *matrix-related bias* implying the influence of the milieu of the analyte that is different from the native samples intended to be measured by MP (83). However, the source of influence may include differences between the analyte, intended to be measured, and measurand itself (e.g. ditauro bilirubin in processed samples vs. conjugated bilirubin in native patient samples, enzymes of non-human origin used to spike the control material). Therefore, the term commutability includes all the differences in MP observed with processed samples, originating from a non-native form of the analyte or by the matrix itself. It has to be taken into consideration that measurands have to be clearly defined when assessing commutability. For example, the same protein can be measured using different immunochemical MPs targeting at different epitopes, thus implying different measurand for the same analyte. The specificity of measurement procedure towards the measurand is an important issue in commutability assessment, and MPs found to be non-specific towards measurand in patient samples are more likely to be the source of noncommutability of RMs. Furthermore, if the origin of differences observed in measurement results is clearly attributed to the influence of an endogenous substance present in abnormal concentration (like high bilirubin concentration in samples), such difference is generally

considered as interference, which magnitude can be further quantified in terms of the analyte and interfering substance (84).

1.2.2 Commutability in EQA programs

Following traceability scheme presented in Figure 1, the critical step in the attempt of standardisation and harmonisation of measurement results is the use of commutable secondary calibrator for value assignment to MPs designed for routine use with native patient samples in medical biochemical laboratories. The true value is assigned by the reference measurement procedure, preferably listed in the JCTLM database. If commutability of RMs used as common calibrators cannot be assured, then comparability, or harmonisation of MPs cannot be expected. The clear example of non-harmonisation due to the noncommutability of RM was described by Zengers et al. (81), on the example of observed differences in EQA results for ceruloplasmin between commonly used nephelometric and turbidimetric methods. All methods were traceable to RM ERM-DA470, certified as a common calibrator for 15 serum proteins, including ceruloplasmin. Although the use of the common calibrator for serum proteins resulted in the reduction of biases between methods for the majority of certified proteins, the results of ceruloplasmin showed large discrepancies between some commonly used methods. It was further investigated and proved that the ERM-DA470 was noncommutable for several method combinations, which resulted in large differences between ceruloplasmin measurements using these methods. The assumption on commutability can even lead to wrong conclusions on standardisation and applicability of MPs for patient samples, leading to even larger bias between methods. For example, Thienpont LM et al. (68) used 14 fresh-frozen, single donation sera to access the trueness of photometric methods for cholesterol and glucose measurement. They found that the mean biases (+5,2% for a cholesterol-oxidase method and +3,7% for glucose-oxidase method) were much higher than almost bias-free results observed in the EQA program using lyophilised samples. Li et al. (85), reported the false sense of confidence in measurement results of GGT coming from one instrument: the results obtained on lyophilised EQA samples were comparable to other used instruments, whereas the results on patient samples revealed the relative difference between samples from 18% to 27%. Further inspection of the differences revealed that the EQA samples were not commutable for this instrument, and thus cannot be compared to a target value and cannot be considered a substitute for patient

samples. In addition, calibration with noncommutable RM may even cause non-pathological results to change to pathological, and *vice versa* (68,72). Although the impact of noncommutable RMs on measurement results is well documented, the assessment of commutability is still not regularly performed and many RMs lack the information on commutability (44,76). Meng et al. (86) examined the commutability of ten commercial control materials used worldwide for triglyceride measurements and discovered that all of the materials showed noncommutability (both positive and negative bias) in 9 out of 14 methods investigated and used in Chinese laboratories.

The commutability of EQA samples is crucial if results from different MPs are to be compared in the same groups and according to the true value of the analyte. In the traceability era, it is EQA samples that serve as post-market vigilance tool for different products used in medical biochemical laboratories and are very often the unique proof to verify the appropriateness of manufacturers' claims in MP (23). EQA monitoring showed on several occasions that even despite clear regulations towards standardisation and traceability, measurement results in native sera show inadequate standardisation and harmonisation even for most common analytes (42,68,87). The role is also educational, because the root-cause of observed bias has to be closely inspected, all stakeholders informed, and possible solutions suggested to manufacturers, regulation bodies and end users. As an example, Figure 4 presents the results for EQA evaluation of trueness of serum alkaline phosphatase (AP) measurement on fresh-frozen serum samples in a group of Italian laboratories, where authors clearly identify the source of recorded discrepancies in EQA results (88). Comparing the results from seven major instrument groups coming from the four manufacturers, they observed clear underestimation on Cobas systems (Roche Diagnostics) and overestimation of AP measurements on AU systems (Beckman Coulter), both being outside of desirable bias for the clinical suitability of the results. After collecting the materials and information on traceability and uncertainty of calibrators from the manufacturers, they found that the Roche systems use an outdated method on their instruments, and Beckman Coulter states the traceability to an internal "master" calibrator, without traceability anchorage to higher-level RMs. Despite to recommended standardisation approach and availability of the IFCC reference measurement procedure, both manufacturers fail to prove compliance with recommendations, which at the end results in poor harmonisation of measurement results for AP between laboratories.

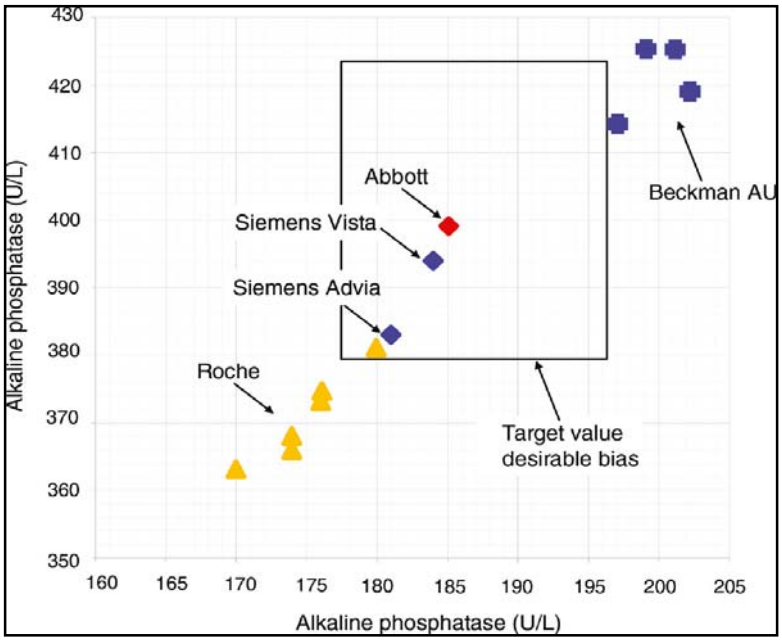


Figure 4. The alkaline phosphatase results for two EQA samples obtained by participants using different measuring systems shown with different colors in a Youden plot.
(Reprinted with permission from reference 88.)

If commutability of EQA samples is not assured or accessed, the participating laboratories in EQA program cannot be evaluated according to unique target value because the difference observed from target value can also be attributed to noncommutability of control material. It is not possible to determine whether any observed biases are caused by inadequate, or noncommutable EQA samples, or genuine biases of evaluated methods. Such evaluation is restricted to forming homogenous peer-groups of participants, usually gathered on the bases of the manufacturer of reagents and instruments used. Peer-groups are expected to have the same matrix-related biases for a given EQA sample, and the evaluation is restricted to the peer-related consensus target value. Such evaluation assures participating laboratories that they use MPs according to manufacturer's specifications, and in agreement with other laboratories using the same technology (52). Peer-group evaluation within EQA is still a necessity for analytes without defined higher-order RM or method, such as lipoproteins, many hormones, tumour markers, etc. Although EQA programs strive to use commutable EQA samples, peer-group evaluation due to potential noncommutability of control material is still used by the majority of providers (22,42,89).

The EQA programs are nowadays classified into 6 categories, according to evaluation capabilities which are dependent on commutability of RMs, target value assignment by

reference laboratory and the use of repeated samples in order to separate differences from bias and/or imprecision of methods (Table 2) (52,90). On the top of the classification is category 1 EQA program with replicate commutable samples in one EQA survey with target values assigned by the higher order reference method. It offers the possibility for evaluation of both laboratories and MPs in medical biochemical laboratories, thus both standardisation achievements and individual laboratory performance EQA programs in categories 3 and 4 also use commutable samples, but have no value assignment by reference MPs, often due to the lack of formally recognised reference systems. Nevertheless, they provide valuable information on harmonisation status of laboratory measurements. Last two categories have samples that are most likely noncommutable and are therefore restricted to peer-group evaluation without being able to further inform participants on standardisation or harmonisation of MPs.

Table 2. Evaluation capability of EQA related to the program design. (Reprinted with permission from reference 52.)

	Evaluation capability										
				Accuracy Individual laboratory					Standardisation or harmonisation		
	Sample characteristics			Relative to participant results			Reproducibility		MP calibration traceability		
Category	Commutable	Value assigned with RMP* or CRM	Replicate samples in survey	Absolute vs RMP or CRM	Overall	Peer group	Individual laboratory interlab CV	MP interlab CV	Absolute vs RMP or CRM	Relative to participant result
1	Yes	Yes	Yes	x	x	x	x	x	x	x
2	Yes	Yes	No	x	x	x		x	x	x
3	Yes	No	Yes		x	x	x	x		x
4	Yes	No	No		x	x		x		x
5	No	No	Yes			x	x	x		
6	No	No	No			x		x		

*RMP- reference measurement procedure, CRM – certified reference material, MP – measurement procedure

1.2.3 Methods for commutability assessment

Different approaches are used for assessing the commutability of RMs. The aim is to provide an objective evaluation of numeric relationship for measurement results of examined measurand in native patient samples and RMs. The approaches differ in the statistical analysis used to describe the relationship, the RM under study (calibrator or control), the number of methods for which commutability has to be assessed and the availability of reference MP for a given measurand.

Describing and evaluating the relationship between patient samples and control materials was initially performed using correspondence analysis (91). It is a multivariate descriptive technique comparing relationships, or associations between studied elements (e.g. patient samples and methods), plotted in the two-dimensional graphs. It provides a "snapshot" of all the data in graphic plots, giving information on the strength of relationships between elements, enabling evaluation of superimposed associations of control materials (92,93). However, it doesn't provide clear numerical criteria in distinguishing commutable from noncommutable materials.

The least-squares linear regression analysis in assessing commutability was proposed by Eckfeldt et al. (94) and it is the most used method in validating commutability of RMs. The protocol was initially used by *College of American Pathologists* (CAP) for control samples and was further adopted and refined in a guideline EP-14 of the *Clinical and Laboratory Standards Institute* (CLSI) (83). In this approach, the relationship between two MPs is obtained with patient samples using regression analysis and two-sided 95% prediction interval for future observations. Measurement results of RMs are further compared to the regression line and its prediction interval. Measurements that fall into limits of 95% prediction interval defined with patient samples are considered commutable whereas the measurements outside the limits are defined as noncommutable (Figure 5). The regression analysis offers an objective, numeric relationship between measurements of patient samples and processed, control samples using two MPs.

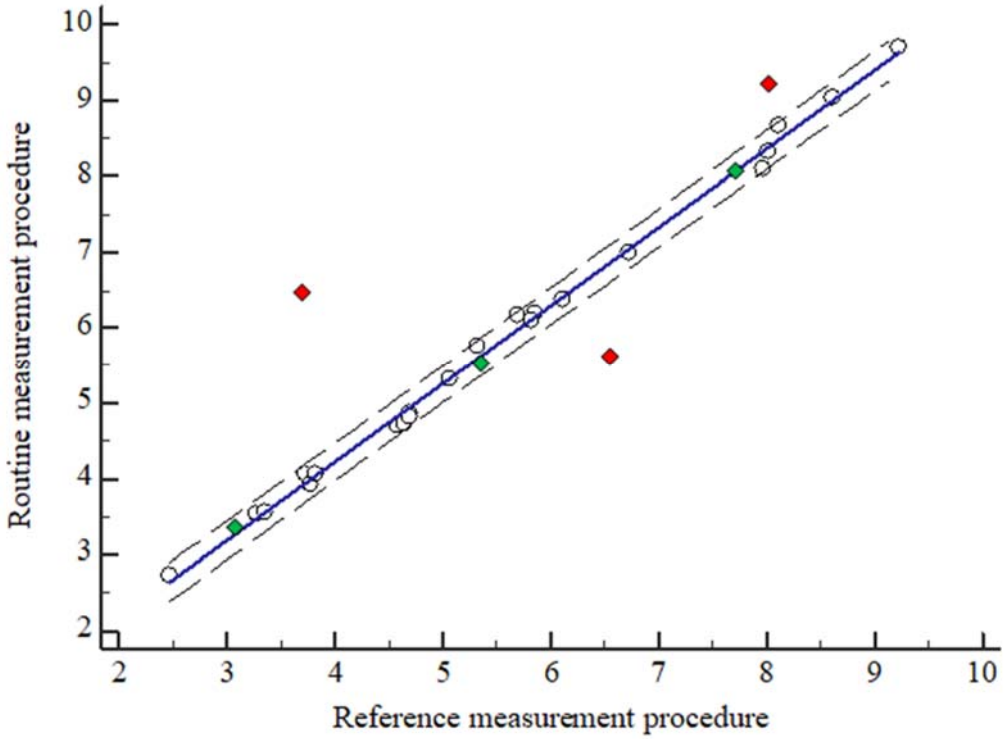


Figure 5. Scatter plots of measurement results of patient samples (black circles) and processed materials (diamonds) on reference and routine MPs. The blue solid line is regression line and black dashed lines present two-tailed 95% prediction interval defined by measurements of patient sera with both MPs. The processed materials falling outside 95% prediction interval are considered noncommutable (red diamonds) and materials inside these limits are commutable (green diamonds).

Initially, ordinary linear regression (ORL) was proposed for analysis. This protocol assumed no variability in comparative method represented on the x-axis and was thus most appropriate for evaluating field methods with reference methods with negligible bias. Such analysis has drawbacks for assessing commutability of EQA samples because numerous methods used in medical biochemical laboratories cannot be considered uncertainty-free, and the conclusion on commutability might theoretically depend on the choice of corresponding axes for each method. The ORL was displaced by Deming regression by some authors and in the third edition or the CLSI document (95) due to the advantage of allowing variability of results for both x and y-axes. In cases where the linear relationship between measurements with two methods cannot be assured, CLSI protocol and some authors suggest the use of best fitting polynomial regression model, with its prediction interval in validating commutability (76,83,93).

Following regression analysis, evaluation of normalised residuals was introduced by Franzini et al. (96) for assessing commutability of control materials. In this analysis, the regression line for two MPs is constructed using patient samples, and the distance of measurement results of RM from the regression line is calculated. The residuals are therefore the differences between the observed and predicted values from the regression analysis. Normalised residuals are calculated by dividing the difference with residual standard deviation (SD_{yx}) of patient sera. RM is considered commutable if its normalised residual is within ± 3 SD_{yx}, as presented in Figure 6. This protocol was used in commutability studies for many RMs and it was noted that it is sensitive to differences in the imprecision of MPs compared, where larger imprecision would cause wider 95% prediction interval and thus more materials to appear commutable (72,97,98). It was suggested that the effect of imprecision can be somewhat reduced using mean values of multiple replicate measurements in the analysis. Having to deal with numerous methods involved in measuring HDL cholesterol in an EQA program, Baadenhuijsen et. al. (99) described an alternate study in order to simplify the native serum acquisition needed for regression analysis (99). This so-called *twin-study* design was a multicentre protocol with the same patient samples (split-patient-sample) being shared between laboratories organized in pairs. The pairs of laboratories were formed to achieve adequate replication and coverage of all methods used in the EQA program. Due to the absence of unbiased reference method for HDL cholesterol measurement, the authors used bivariate regression analysis according to Passing and Bablok (100). It is a robust, distribution-free method that is not sensitive to outliers, does not require constant standard deviation over the measuring range and assumes variability in both methods under study (101). However, the prediction intervals are larger than those coming from the procedures based on least-squares linear regression, which may result in more

accepted control materials for commutability than an analysis based on least-squares linear regression. Adding to a larger confidence interval using distribution-free regression analysis, the scatter of results coming from laboratory pairs is larger, which has been seen by the authors as an advantage since imprecision of methods and potential matrix-effect are presented to the maximum degree. To minimize the effects of larger observed imprecision, the perpendicular distances of RMs were normalized by expressing them as multiples of the state-of-the-art within-laboratory SD observed in an EQA program. Using the same criteria of ± 3SD being acceptable (commutable), the authors were able to evaluate commutability of RM according to state-of-the-art criteria of their own EQA program. Once established, the commutability is further monitored using native *spy sample* with approximately the same analyte concentration. The ratio between results obtained with EQA sample and the native sample is compared and the significance of differences examined using a Student's *t*-test (23).

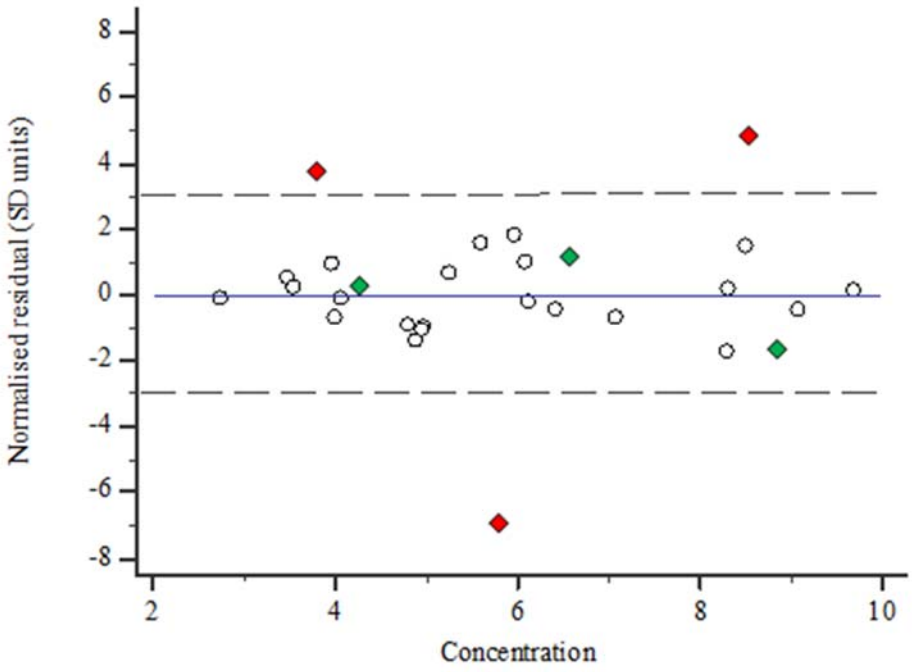


Figure 6. Commutability assessment of RM (diamonds) using normalized residuals (circles) and ±3 SD limits (dashed black line). Noncommutable RMs are presented as red diamonds and commutable RM as green diamonds.

All these analysis models adopted statistical limits to validate commutability of RMs; using boundaries of 95% prediction interval or limits defined by a number of normalised residuals

from the regression analysis. In the approach from Ricos et al. (102) the RM residuals were expressed as percentage bias from predicted values and further compared by the biological variation-based criteria for bias. In addition, the authors compared three criteria in assessing commutability of RM in creatinine analysis: the 95% prediction interval boundaries, ± 2 standardised residual criteria from Passing and Bablok regression and comparison of percentage bias observed to fixed limits of bias. It was concluded that at high concentration levels, all three models gave concordant results, whereas at normal and low concentrations, ±2 standardised residual criteria were too permissive classifying more RM as being commutable. The observation was explained by non-constant variability along measuring range where larger variability can be seen with low concentration levels.

The difference in bias approach in the evaluation of EQA samples for measurements of HDL and LDL cholesterol was further investigated by two independent groups of authors (103,104). In both groups bias of measurements of patient samples and control samples with the associated uncertainty of measurements was compared to fixed criteria of allowed bias from CDC's (*Centers for Disease Control*) *Lipid Standardisation Program*, considered as medical requirement criteria. EQA samples validated appeared to be mostly noncommutable when using favourable medical requirement criteria over criteria based on random error. Further discussed, the approach offers evaluation of RM according to clinical intended use, but the criteria seem to be too stringent considering the fact that if patient samples (commutable by definition) were evaluated according to the same criteria, only 23% - 27% were found to be commutable, against 83% - 87% using criteria based on random error components (104). The authors explain that the possible explanation lies in the specimen specific effects known to be influencing homogenous methods for HDL and LDL and the performance characteristics of MPs under evaluation.

The assessment of commutability using fixed criteria was very recently proposed by *IFCC Working Group on Commutability* (IFCC-WGC) (105–107). The recommendations are divided into three parts in order to cover many aspects of commutability: definitions and descriptions of RMs for which commutability assessment should be used, the experimental design, requirements for clinical samples and MPs included in design, evaluation criteria to determine commutability for various RMs and the statistical approaches in validating commutability of EQA samples and calibrators. The IFCC-WGC describes statistical criteria in evaluating commutability as less desirable and does not recommend such criteria, stressing the importance of applying equal limits for the same measurand using different MPs. This was recognised as particularly important when comparing results of the RM on MPs with different precision

profiles, where less precise methods would yield more materials to be commutable comparing to the comparison of high-precision methods with consequent narrow confidence intervals. The authors even suggest the initial assessment or precision profiles for individual MPs to verify their appropriateness, or *fitness-for-purpose* in commutability evaluation protocol described. Besides fixed commutability criterion for assessment of RMs and identification of precision of MPs as an important factor influencing commutability outcome, the recommendations use the separate experimental design for different RMs, i.e., calibrators and control samples. The authors recommend that commutability criteria be chosen according to the intended use of RM; being expressed as a fraction of uncertainty needed for calibrators to be used in traceability hierarchy producing allowable bias in clinical samples or expressed as a fraction of bias component of the APS in EQA control samples evaluation.

Experimental design for assessing commutability of control samples includes measuring clinical samples and control samples using all MPs included in commutability assessment. The difference in bias between an RM and average bias of clinical samples is determined, the uncertainty of that difference calculated (and multiplied by suitable coverage factor, usually 1.96 for 95% level of significance), and compared to previously established "allowable bias" or commutability criterion range. Thus, an important part of commutability assessment is not only the average difference in bias observed for RM and clinical samples, but also the uncertainty of that bias, which has to fit in the commutability criterion for the control sample to be considered commutable. The uncertainty in bias has two components: uncertainty of the estimated bias for clinical samples and uncertainty of the estimated bias for RM, resulting in total uncertainty, or error bars (Figure 7) around the average difference of RM and clinical samples. In order to be able to estimate these uncertainties, evaluate precision profiles and sample-specific effects for MPs under study, assuring constant scatter across the concentration interval, at least 30 clinical samples should be measured in triplicate measurements. The uncertainty of estimated bias from clinical samples is calculated using pooled standard deviations from replicate measurements, after checking that the bias change from consecutive measurements is relatively small. If the constant width of the scatter cannot be observed, the transformation of the data should be used to assure approximately constant bias along the concentration range. The uncertainty of difference in bias for RM consists of pooled standard deviations of replicate measurements (at least three) and position effects (at least five). Because the random effects may have a significant influence on commutability decision, the IFCC-WGC suggests that methods should be evaluated and pre-qualified for commutability assessment experiment, where only methods with adequate precision should be used.

Figure 7 presents the example of commutability assessment recommended by IFCC-WGC on two combinations of MPs using fixed criteria for commutability assessment. Due to the fact that the difference in bias was not constant over the concentration range, the data were transformed to ln(concentration) to give a constant scatter.

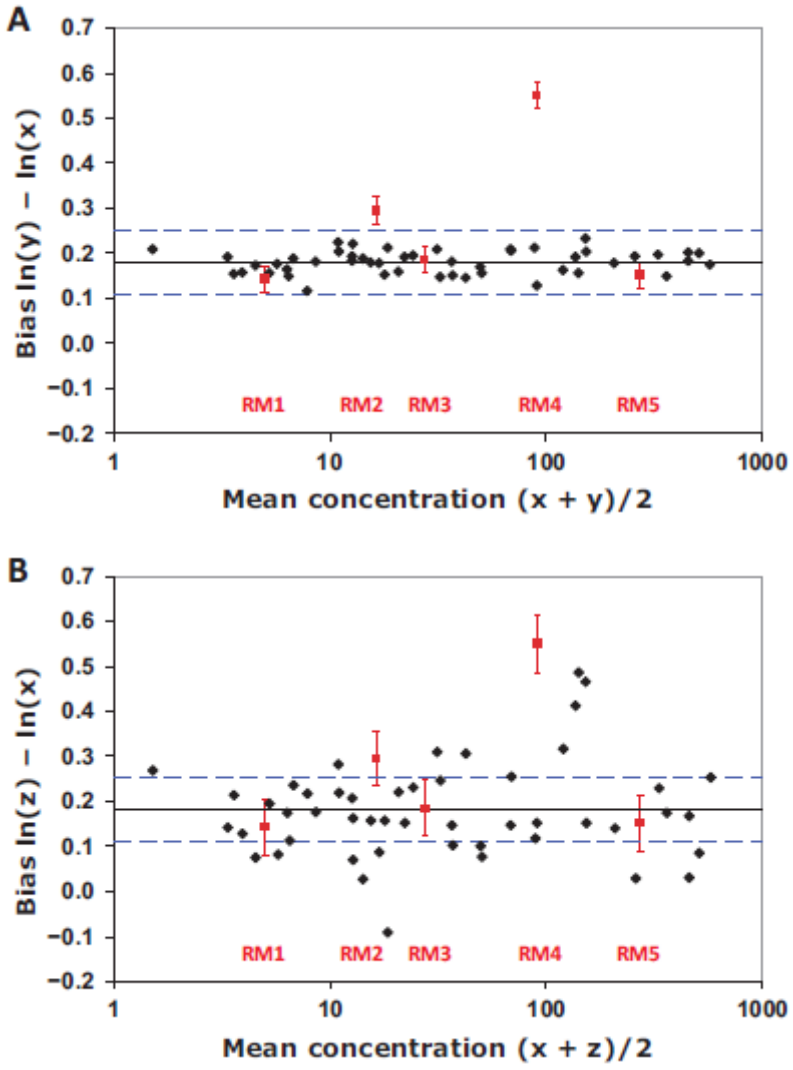


Figure 7. An example of commutability assessment based on the difference in bias between results for clinical samples (black diamonds) and 5 RMs (red squares) of MPs y and x (A) and z and x (B). The ordinates of the two graphs show the biases for logarithmic (ln) transformation of concentration compared to the mean concentration of samples (on the abscissa) on two measurements procedures. (Reprinted with permission from reference 106.)

The results for methods y and x (Figure 7A) show small random error (satisfactory precision) and sample-specific influences whereas the results for the method z and x (Figure 7B) are more

scattered, suggesting less precision for method z and thus wider uncertainty limits of observed bias. RM1, RM3 and RM5 are commutable and RM2 and RM4 are noncommutable for the method combinations y and x. Due to the larger random effects and thus larger uncertainty of observed bias, only RM3 is commutable for method combinations x and z. Commutability of RM1, RM2, and RM5 remains inconclusive because the error bars of those materials span outside the fixed commutability criterion.

IFCC-WGC recommends the assessment of commutability of calibrators by means of their ability to serve as a tool for successful harmonisation of clinical samples' measurement results using different MPs. Although random and sample-specific effects between MPs can cause different results for clinical samples, the cause of the difference can also be the bias between MPs. The causes for bias are all connected to calibration procedure, and possible sources are an inappropriate model for the calibration curve, incorrect values of the calibrators, and a difference in behavior between calibrator and clinical samples in MPs (different response for the same concentration), or noncommutability of the calibrator for those MPs. The bias caused by calibration with noncommutable calibrator can be reduced with the use of same, commutable calibrator for all MP used for measuring clinical samples. As the IFCC-WGC recommends, after initial evaluation of between-measurements differences for clinical samples, the recalibration with the calibrator under evaluation for commutability is performed, and the resultant differences between means for the methods are compared to previously established commutability criterion. If the observed differences are significant after the recalibration in a way to fit-in to allowed bias between methods, the calibrator is considered commutable. If such reduction in bias cannot be observed, the commutability cannot be confirmed, and other sources of calibration bias must be investigated prior to concluding on its commutability, such as high imprecision of the method, a poor fitting mathematical model for the calibration curve, individual sample-specific interferences and others. Figure 8 shows the recalibration effects of evaluated calibrator between 7 MPs. The between-methods differences for clinical samples are significantly reduced after recalibration of all MPs except for the MP6. Since the differences for the clinical samples measured using MP 1-5 and MP7 after recalibration falls into commutability criterion of $\leq$ 6%, the calibrator is considered commutable for those MPs. The commutability of calibrator MP6 cannot be confirmed and the manufacturer should be notified of such a conclusion.

The analysis of commutability according to fixed, previously established criteria according to the intended use of RM, seem to provide an objective assessment of commutability in various MPs. Using such criteria, commutability of the control samples should be assessed using a

commutability criterion that would be only a fraction of APS in the EQA scheme, although this fraction remains undefined. Furthermore, the strict prerequisites for adequate precision of methods to be evaluated potentially leave out many MPs used by laboratories. In addition, the random effects observed for clinical samples may still be very different for MPs under evaluation and yield larger uncertainties of the observed bias causing more materials to appear inconclusive or noncommutable.

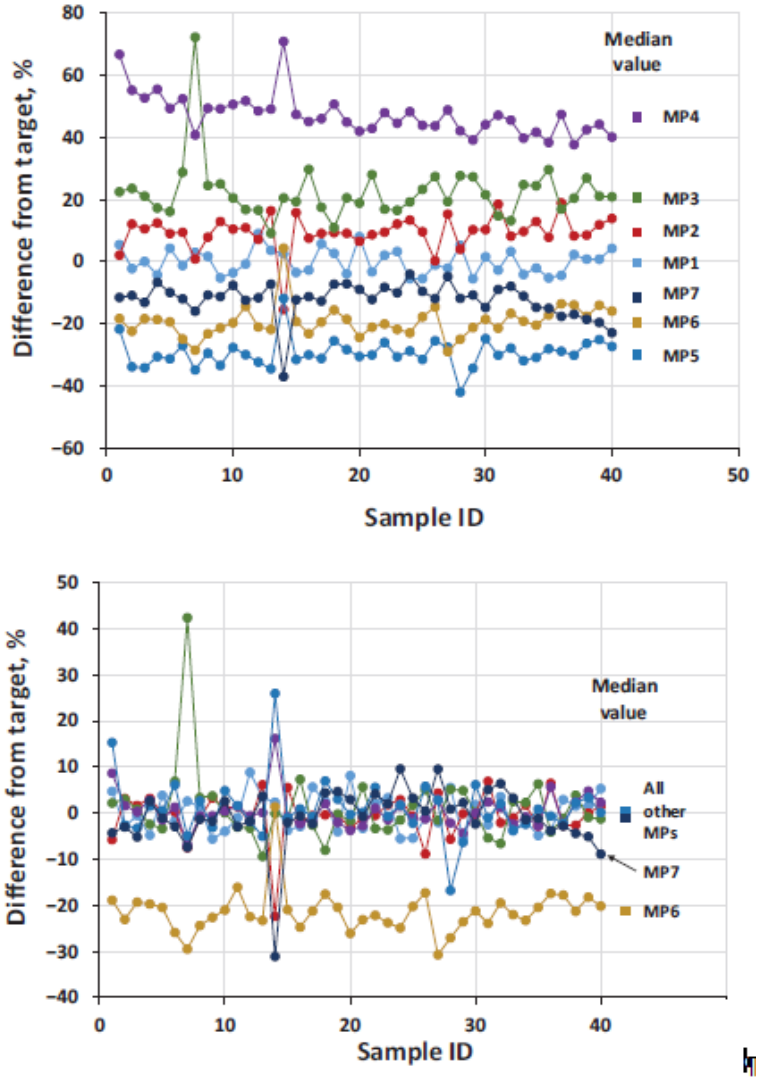


Figure 8. Difference in percent from the target value (trimmed mean) for 40 clinical specimens from 7 MPs prior to recalibration (top graph) and after recalibration (bottom graph) with evaluated calibrator. The color of each dot is representative for the corresponding measurement procedure. Sample ID – Sample identification, MP 1-7 – measurement procedure 1-7. (Reprinted with permission from reference 107.)

Since the recommendations from the IFCC-WGC were just recently published, there are still no published data on the application (or use) of fixed criteria in the assessment of commutability of control samples used in EQA. It remains to be seen whether demanding economic and logistic experiment design will yield the expected benefit for both participant laboratories and EQA providers in evaluating the control materials to be used for interlaboratory comparison and trueness assessment.

2. AIM OF RESEARCH

Aims of this research are:

- Assessment of commutability of EQA control samples for most common biochemical tests measured in medical biochemical laboratories using statistical models for comparison and evaluation of the significance of observed differences between measurements obtained on serum samples and lyophilised control samples, both analysed in an EQA setting.
- Establishment of the new commutability evaluation approach, i.e. *false flagging method:*
    - Establishment of commutability limits as a maximum allowable rate of falsely flagging laboratories and MPs based on the results obtained on serum and control samples used in the EQA program;
    - Evaluation of commutability limits for control samples using APS criteria and intended use of control samples for assessment of laboratory and MP's performance;
    - Validation of commutability limits on EQA results of CROQALM for most common biochemical tests (ALT, AP, AMY, AST, calcium, chloride, total cholesterol, CK, creatinine, GGT, glucose, HDL cholesterol, iron, LDH, phosphate, potassium, proteins, sodium, bilirubin, triglycerides, urate and urea).
- Evaluation of commutability according to regression analysis recommended by widely used CLSI document EP14 for glucose, cholesterol, triglycerides, HDL cholesterol, urea, creatinine, sodium, potassium, chloride, AST, ALT and GGT on some of the most used instruments in CROQALM.
- Comparison of the regression analysis (CLSI document EP14) with the proposed false flagging method in commutability evaluation of EQA control samples.
- Assessment of advantages and disadvantages of the newly proposed false flagging method for commutability testing of control samples in the EQA program.

3. MATERIALS AND METHODS

3.1 Materials

Materials used in commutability evaluation include native, off-the-clot serum samples from voluntary donors, native serum samples from voluntary donors spiked with glucose, urea, sodium, potassium, chloride, bilirubin, copper and residual patient serum samples collected after routine analysis in the medical biochemical laboratory.

3.1.1 Native serum samples

Blood was collected at the *Croatian Institute for Transfusion Medicine*, Zagreb, Croatia, from voluntary donors. In order to be eligible for blood donation and this study, all blood donors had to meet the mandatory criteria stated in the Law on Blood and Blood Components (108), and no other additional criteria were required. All voluntary donors agreed and signed the informed consent prior to donation. A volume of 450 ml of blood was collected under sterile conditions in plastic bags without anticoagulant added and later used as starting material for EQA native serum samples. After 2-3 hours of spontaneous clotting, the blood was centrifuged, and off-the-clot serum collected in a second plastic bag. Centrifugation and serum collecting step was repeated to eliminate visible fibrin and residual cells from the material. The serum is tested and found negative for HCV RNA, HIV 1/2 HBV DNA, HIV Ag, anti-HIV 1/2, anti HCV, HBsAg and anti TP. The yield of the serum was about 170-190 mL, depending on the dose and clotting time. Native serum from two donors was mixed in a sterile plastic bag for one hour. The serum is further aliquoted in 205 sterile plastic tubes and stored at +4^{o}C prior to shipment.

3.1.2 Spiked serum samples

To achieve a high level of particular measurands, appropriate amounts of native serum in the second and third EQA surveys were spiked with the following solutions:

- Glucose solution (1 M), prepared by dissolving 18.02 g of D-(+)-glucose anhydrous (Claro-Prom, Zagreb, Croatia) in sterile deionised water (100 mL stock solution).
- Urea solution (1 M), prepared by dissolving 6.0 g of urea (Merck KGaA, Darmstadt, Germany) in sterile deionised water (100 mL stock solution).
- Solution of NaCl (1 M), prepared by dissolving 5.84 g NaCl (Merck KGaA, Darmstadt, Germany) in sterile deionised water (100 mL stock solution).
- Bilirubin solution (6.3 mM), prepared by dissolving 0.37 g bilirubin (Merck KGaA, Darmstadt, Germany) in the mixture of 2.0 mL 0,1M Na_2CO_3 (Merck KGaA, Darmstadt, Germany) and 1.5 mL 0,1 M NaOH (Merck KGaA, Darmstadt, Germany) and then (subsequently, then) reconstituted in sterile deionised water (100 mL stock solution). The stock solution was stored in dark, protected from light.
- Conjugated bilirubin solution (2.85 mM, 5 mL), prepared by dissolving 12.0 mg of bilirubin conjugate, ditaurate, disodium salt (Merck KGaA, Darmstadt, Germany) in 5.0 mL of sterile deionised water. The stock solution was stored in dark, protected from light.
- Magnesium standard (41.1 mM) (Perkin Elmer, Waltham, MA, USA).
- Cu standard (15.74 mM) (Perkin Elmer, Waltham, MA, USA).
- KCl, infusion concentrate (1M) (Croatian Institute for Transfusion Medicine, Zagreb, Croatia)

Spiking solution for the second EQA survey. The solution is prepared from stock solutions by mixing 2.0 mL 1 M glucose, 2.0 mL 1M urea, 4.0 mL 1M NaCl, 8 mL 6.3 mM bilirubin, 2.0 mL Mg, and 0.2 mL Cu solutions (total volume 18.2 mL).

EQA samples for the second survey were prepared by adding 18.2 mL of spiking solution to 200.0 mL of previously prepared native serum. Spiked serum was further mixed for 2 hours, aliquoted in 205 plastic sterile tubes and stored at +4°C prior to shipment.

Spiking solution for the third EQA survey. The solution is prepared by mixing 1.2 mL 1 M glucose, 1.2 mL 1M urea, 1.5 mL 1M NaCl, 2 mL 2.85 mM conjugated bilirubin, 1.2 mL Mg, and 0.1 mL KCl solutions (total volume 7.2 mL).

Serum samples for the third EQA survey were prepared from single donor blood. After preparation, native serum was split into two volumes: 105.0 mL (V1) and 95.0 mL (V2). V1 was ready to use (native serum sample) after aliquoting in 190 sterile plastic tubes. V2 was spiked with spiking solution for the third survey, and then mixed for two hours, aliquoted in190 plastic tubes and stored at +4^{o}C prior to shipment (spiked serum samples).

3.1.3 Residual patient serum samples

Residual patient serum samples were collected after routine analysis in the *Department of Laboratory Diagnostics*, General Hospital Pula, Croatia. The samples were collected from patients which signed the informed consent on the use of the leftover material after routine analysis. The blood was drawn from the antecubital vein in plastic serum tubes without anticoagulant used. The samples were selected in a manner to meet concentration ranges needed to be evaluated in the CLSI protocol for commutability evaluation.

3.1.4 Lyophilised commercial control samples

Three lyophilised, human-based control samples from two manufacturers were used. The controls were named C1/2016, C2/2016 and C3/2016, according to the use in appropriate EQA surveys (1-3):

C1/2016 (EQA survey 1): Seronorm^{TM} Human, LOT 1412548 (SERO, Billingstad, Norway);

C2/2016 (EQA survey 2): Seronorm^{TM} Human High, LOT 1403083 (SERO, Billingstad, Norway);

C3/2016 (EQA survey 3): Human Assayed Control – Level 1, LOT HSN026 (Fortress Diagnostics, Antrim, UK).

Control materials Seronorm™ Human from SERO (C1/2016 and C2/2016) are claimed to be "excellent choice for laboratories seeking a commutable material for both precision and accuracy monitoring", whereas control material Human Assayed Control from Fortress diagnostics has no claims on commutability.

Lyophilised control samples were distributed in the original vials. The material was dissolved in 5.0 mL of sterile deionised water (with occasional gentle mixing by inverting the vial several times) by participating laboratories, following written instructions. After 30 minutes, the samples were ready for analysis.

3.2 Procedure for commutability evaluation of control samples using regression analysis

Commutability evaluation of lyophilised commercial control samples was performed on three occasions (December 2016, May 2017 and October 2017), according to the protocol recommended by CLSI guideline EP14-A3 (95). The evaluation was performed on five instruments: Roche Cobas 6000 c501 (Roche Diagnostics, Mannheim, Germany), Roche Cobas Integra 400 plus (Roche Diagnostics, Mannheim, Germany), Abbott Architect c4000 (Abbot Laboratories, Chicago, IL, USA), Beckman Coulter AU 680 (Beckman Coulter, Brea, CA, USA) and Siemens Dimension Xpand (Siemens Healthineers, Newark, DE, USA). Routine methods were used for analysis of 12 analytes: glucose (Hexokinase method), total cholesterol (Cholesterol oxidase/peroxidase – phenol/4-aminophenazone method), triglycerides (Glycerol phosphate oxidase/peroxidase - phenol/4-aminophenazone method), HDL cholesterol (Homogeneous enzymatic method), urea (Urease/Glutamate dehydrogenase, method), creatinine (Compensated Jaffe method), sodium (Indirect ISE method), potassium (Indirect ISE method), chloride (Indirect ISE method), alanine aminotransferase (IFCC method), aspartate aminotransferase (IFCC method), and gamma glutamyltransferase (IFCC method). The instruments chosen for assessment are the ones that have the largest number of participants in CROQALM scheme and are mostly homogeneous systems where both instruments and reagents come from the same manufacturer.

The 20 - 22 residual serum samples for each analyte group (glucose and lipids group; urea, creatinine and electrolytes group; enzymes group) were sent wrapped in cooled packages together with control samples to four laboratories participating in CLSI study of commutability. All samples were transported and analysed within 24 hours of collection. For each sample and analyte, the analysis was performed in triplicate measurements and the average of those is used for further calculations.

The CLSI protocol was performed on three occasions to be able to assure collecting fresh patient samples that would span the broad analytical range covering low, normal and high levels of each analyte. Besides concentration levels, the choice of the residual patient samples was mostly dependent on the residual volume left after routine analysis in the laboratory and absence of any known or visible interferences (for example haemolysis, icteria and lipemia). Due to the lack of reference MP for comparison, statistical analysis of results for each MP (based on instrument and analytical method used) was initially performed using Deming regression analysis, as suggested in the CLSI EP14-A3 guideline (95). The regression line was defined with patient samples, and a 95% prediction interval for the new observations was calculated according to the same recommendations.

Considering the number of results from patient samples that were outside of proposed 95% prediction interval serving as a commutability criterion, the regression analysis was done according to the previous edition of same CLSI guideline (EP-14-A2), using simple linear regression analysis. The control samples whose results exceeded the limits of the 95% prediction interval around the regression line calculated for the patient samples were considered as noncommutable.

3.3 Study design of commutability evaluation of control samples within EQA

The serum samples and control samples were analysed in three scheduled CROQALM surveys in March, June and September 2016. The samples were shipped to participant laboratories at ambient temperature together with written instructions on analysis details. The laboratories were instructed to analyse the samples as soon as possible after receipt, both lyophilised control and serum samples in the same run on the instrument, using the routine MPs

used in the laboratory. The outline of the EQA sample analysis in each survey is shown in Figure 9.

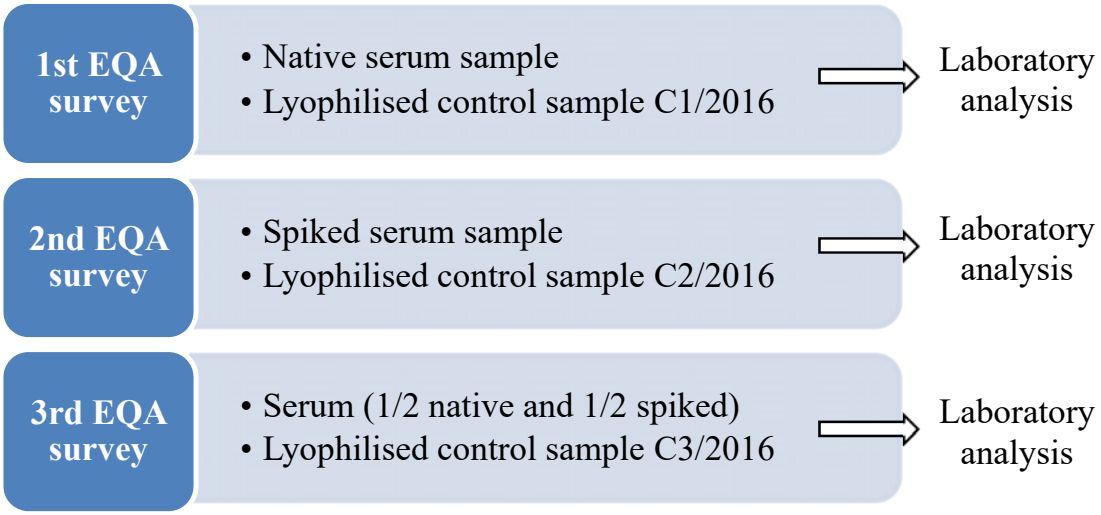


Figure 9. The course of sample analysis within CROQALM

The number of participating laboratories in each survey varied from 180 in survey 1, 182 in survey 2, and 184 in survey 3, depending on the laboratories' voluntary participation in the EQA study surveys. The majority of laboratories received the samples one day after shipment (surveys 1-3: 170/180, 170/182, 169/184, respectively) and analysed the samples promptly upon receipt. After analysis, the laboratories entered the results through the web interface of inlab2*QALM software for quality evaluation in laboratory medicine (IN2 Group Ltd., Zagreb, Croatia). The laboratories chose the method, instrument and reagent that they used for analysing the samples.

The participants were instructed to measure all the analytes from the biochemistry module of CROQALM which includes 32 parameters, if those are in the scope of the laboratory's routine operation. In order to form homogeneous peer groups based on MPs used for analysis, the results for the same method and instrument used for each analyte were grouped together. The number of data received for analytes that were too few to include at least two MPs to be compared across three EQA surveys, were not included in the study. The results for 22 analytes were included in further statistical analysis: alanine aminotransferase (ALT), alkaline phosphatase (AP), alpha amylase (AMY), aspartate aminotransferase (AST), total calcium (calcium), chloride, total cholesterol (cholesterol), creatine kinase (CK), creatinine, gamma

glutamyltransferase (GGT), glucose, HDL cholesterol (HDL), iron, lactate dehydrogenase (LDH), inorganic phosphate (phosphate), potassium, sodium, total bilirubin (bilirubin), total protein (proteins), triglyceride, urate and urea.

The results of each MP were tested for outliers using the Grubbs (109) test at a significance level of 95%. Only MPs with 6 and more participants after outlier exclusion were included in final MPs groups for statistical analysis.

Table 3 shows all MPs included in the commutability evaluation of control samples within EQA. Overall, 143 MPs groups were formed based on a different combination of analytical methods and instruments used for measurements of controls and serum samples. Depending on the analyte, 3 to 7 different instruments were used for measurement, and considering the methods applied to each instrument, the number of MPs varied from 3 to 9 for each analyte.

Table 3. MPs included in commutability study based on analytical method and instrument used for each analyte.

Analyte	Method	Instrument	MP
Alanine aminotransferase (ALT)	**IFCC** (37 ºC, TRIS buffer, pH 7,15, L-Alanine, Oxoglutarate, NADH, Lactate dehydrogenase, Pyridoxal phosphate) **Photometry UV** (37 ºC, TRIS buffer, pH 7,15, L-Alanine, Oxoglutarate, NADH, Lactate dehydrogenase)	Abbott Architect c Beckman Coulter AU Roche Cobas c Roche Cobas Integra Roche Hitachi Siemens Dimension	IFCC- BC AU IFCC-SD Photometry UV-BC AU Photometry UV-AA Photometry UV-RCI Photometry UV-RCc Photometry UV-RH
Alkaline phosphatase (AP)	**IFCC** (37 ºC, 2-Amino-2-methyl-1-propanol, pH 7,2, 4-Nitrophenyl phosphate, Zn^{2+}, Mg^{2+}, HEDTA)	Abbott Architect c Beckman Coulter AU Horiba Pentra Roche Cobas c Roche Cobas Integra Roche Hitachi Siemens Dimension	IFCC- AA IFCC- BC AU IFCC- HP IFCC- RCc IFCC- RCI IFCC- RH IFCC- SD
Alpha-amylase (AMY)	**IFCC** (37 ºC, HEPES, pH 7,0, 4,6-Ethylidene(G1)-4-nytrophenyl (G7)-2-maltoheptaoside, Sodium chloride, Calcium chloride, Alpha –glucosidase) **Photometry, CNP-G3**	Abbott Architect c Beckman Coulter AU Roche Cobas c Roche Cobas Integra Roche Hitachi	IFCC- AA IFCC- BC AU IFCC- RCc IFCC- RCI IFCC- RH CNP-G3- SD
Aspartate aminotransferase (AST)	**IFCC** (37 ºC, TRIS buffer, pH 7,65, L-aspartate, oxoglutarate, NADH, malate dehydrogenase, Pyridoxal phosphate) **Photometry UV** (37 ºC, TRIS buffer, pH 7,65, L-aspartate, oxoglutarate, NADH, malate dehydrogenase)	Abbott Architect c Beckman Coulter AU Roche Cobas c Roche Cobas Integra Roche Hitachi Siemens Dimension	IFCC- BC AU IFCC- RH IFCC- SD Photometry UV-BC AU Photometry UV-AA Photometry UV-RCI Photometry UV-RCc Photometry UV-RH
Total calcium (Calcium)	**Photometry, Arsenaso III** **Photometry, NM-BAPTA** **Photometry, cresolphthalein**	Abbott Architect c Beckman Coulter AU Roche Cobas Integra Roche Hitachi Siemens Dimension	Arsenaso III- AA Arsenaso III- BC AU NM-BAPTA- RCI cresolphthalein- BC AU cresolphthalein- RCI cresolphthalein- SD
Chloride	**Indirect ISE**	Abbott Architect c Beckman Coulter AU Siemens Dimension	Indirect ISE- AA Indirect ISE- BC AU Indirect ISE- SD

Analyte	Method	Instrument	MP
Total cholesterol (Cholesterol)	**CHOD-PAP**	Abbott Architect c Beckman Coulter AU Horiba Pentra Roche Cobas c Roche Cobas Integra Roche Hitachi Siemens Dimension	CHOD-PAP- AA CHOD-PAP- BC AU CHOD-PAP- HP CHOD-PAP- RCc CHOD-PAP- RCI CHOD-PAP- RH CHOD-PAP- SD
Creatine kinase (CK)	**IFCC** (37 ºC, Imidazole, pH 6,5, Creatine phosphate, ADP, EDTA, Mg^{2+}, N-aceyl-L-cysteine, AMP, P^1P^5-diAP, NADP, Hexokinase, Glucose-6-phosphate dehydrogenase	Abbott Architect c Beckman Coulter AU Roche Cobas c Roche Cobas Integra Roche Hitachi	IFCC-AA IFCC- BC AU IFCC- RCc IFCC-RCI IFCC-RH
Creatinine	**Compensated Jaffe** (IDMS and NIST SRM 967 traceable) **Enzymatic** (IDMS and NIST SRM 967 traceable) **Non-compensated Jaffe**	Abbott Architect c Beckman Coulter AU Roche Cobas c Roche Cobas Integra Roche Hitachi Siemens Dimension	Compensated Jaffe- AA Compensated Jaffe- BC AU Compensated Jaffe- RCc Compensated Jaffe- RCI Compensated Jaffe- RH Compensated Jaffe- SD Enzymatic- BC AU Non-compensated Jaffe- BC AU Non-compensated Jaffe- RH
Gamma glutamyltransferase (GGT)	**IFCC** (37 ºC, Glycylglycine, pH 7,7, L-y-Glutamyl-3-carboxy-4-nitroanilide)	Abbott Architect c Beckman Coulter AU Horiba Pentra Roche Cobas c Roche Cobas Integra Roche Hitachi Siemens Dimension	IFCC- AA IFCC- BC AU IFCC- HP IFCC-RCc IFCC- RCI IFCC- RH IFCC- SD
Glucose	**GOD-PAP** **Hexokinase**	Abbott Architect c Beckman Coulter AU Horiba Pentra Roche Cobas c Roche Cobas Integra Roche Hitachi Siemens Dimension	GOD-PAP- BC AU GOD-PAP- HP GOD-PAP- RH Hexokinase- AA Hexokinase- BC AU Hexokinase- RCc Hexokinase- RCI Hexokinase- SD
HDL cholesterol (HDL)	**Homogeneous enzymatic**	Abbott Architect c Beckman Coulter AU Horiba Pentra Roche Cobas c Roche Cobas Integra Roche Hitachi Siemens Dimension	Homogeneous- AA Homogeneous- BC AU Homogeneous- HP Homogeneous- RCc Homogeneous- RCI Homogeneous- RH Homogeneous- SD

Analyte	Method	Instrument	MP
Iron	**Photometry, Ferene** **Photometry, Ferrozine** **Photometry, TPTZ**	Abbott Architect c Beckman Coulter AU Horiba Pentra Roche Cobas c Roche Cobas Integra Roche Hitachi Siemens Dimension	Ferene AA Ferene- HP Ferene- RH Ferene- SD Ferrozine- RCc Ferrozine- RCI Ferrozine- RH TPTZ- BC AU
Lactate dehydrogenase (LDH)	**IFCC** (37 ºC , N-Methyl-D-glucamine, L-(+)-Lactate, NAD^+	Abbott Architect c Beckman Coulter AU Roche Cobas c Roche Cobas Integra Roche Hitachi Siemens Dimension	IFCC- AA IFCC- BC AU IFCC- RCc IFCC- RCI IFCC- RH IFCC- SD
Inorganic phosphate (Phosphate)	**Photometry, Ammonium molybdate**	Abbott Architect c Beckman Coulter AU Roche Cobas c Roche Cobas Integra	Ammonium- molybdate- AA Ammonium- molybdate- BC AU Ammonium-molybdate- RCc Ammonium- molybdate- RCI
Potassium	**Flame emission photometry (FES)** **Indirect ISE**	Abbott Architect c Beckman Coulter AU Ciba Corning Roche Cobas c Roche Cobas Integra Siemens Dimension	FES- CC Indirect ISE- AA Indirect ISE- BC AU Indirect ISE- RCc Indirect ISE- RCI Indirect ISE- SD
Sodium	**Flame emission photometry (FES)** **Indirect ISE**	Abbott Architect c Beckman Coulter AU Ciba Corning Roche Cobas c Roche Cobas Integra Siemens Dimension	FES-CC Indirect ISE- AA Indirect ISE- BC AU Indirect ISE- RCc Indirect ISE- RCI Indirect ISE- SD
Total bilirubin (Bilirubin)	**Photometry, Diazo**	Abbott Architect c Beckman Coulter AU Horiba Pentra Roche Cobas c Roche Cobas Integra Roche Hitachi Siemens Dimension	Diazo- AA Diazo- BC AU Diazo- HP Diazo- RCc Diazo- RCI Diazo- RH Diazo- SD
Total proteins (Proteins)	**Photometry, Biuret**	Abbott Architect c Beckman Coulter AU Roche Cobas c Roche Cobas Integra Roche Hitachi Siemens Dimension	Biuret- AA Biuret- BC AU Biuret- RCc Biuret- RCI Biuret- RH Biuret- SD
Triglycerides	**GPO-PAP**	Abbott Architect c Beckman Coulter AU Roche Cobas c Roche Cobas Integra Roche Hitachi Siemens Dimension	GPO-PAP- AA GPO-PAP- BC AU GPO-PAP- RCc GPO-PAP- RCI GPO-PAP- RH GPO-PAP- SD

Analyte	Method	Instrument	MP
Urate	**Uricase UV**	Abbott Architect c	Uricase- BC AU
		Beckman Coulter AU	Uricase- RH
	Uricase/POD	Roche Cobas c	Uricase- SD
		Roche Cobas Integra	Uricase, POD- AA
		Roche Hitachi	Uricase, POD- BC AU
		Siemens Dimension	Uricase, POD- RCc
			Uricase, POD- RCI
			Uricase, POD- RH
Urea	**Urease, GLDH**	Abbott Architect c	Urease, GLDH- AA
		Beckman Coulter AU	Urease, GLDH- BC AU
		Horiba Pentra	Urease, GLDH- HP
		Roche Cobas c	Urease, GLDH- RCc
		Roche Cobas Integra	Urease, GLDH- RCI
		Roche Hitachi	Urease, GLDH- RH
		Siemens Dimension	Urease, GLDH- SD

AA – Abbott Architect c; BC AU - Beckman Coulter AU; CC – Ciba Corning; HP - Horiba Pentra; RCc – Roche Cobas c; RCI - Roche Cobas Integra; RH – Roche Hitachi; SD – Siemens Dimension; PP – Pyridoxal-5'-phosphate; CNP-G3 – 2-chloro-4-nitrophenyl-α-D-maltotrioside, NM-BAPTA - 5-nitro-5'-methyl-(1,2-bis(o-aminophenoxy)ethan-N,N,N',N'-tetraacetic acid; ISE – Ion-selective Electrode; CHOD-PAP – Cholesterol oxidase/peroxidase – phenol/4-aminophenazone; P^1P^5-diAP - P^1P^5-Di(adenosine-5'pentaphosphate; P GOD-PAP – Glucose oxidase/peroxidase- phenol/4-aminophenazone ; TPTZ – 2,4,6,-Tripyridyl-s-triazine; GPO-PAP – Glycerol phosphate oxidase/peroxidase - phenol/4-aminophenazone; POD – Peroxidase; GLDH – Glutamate dehydrogenase

The results received from the analysis of serum samples are each time compared to results received for lyophilized control samples on the same survey.

Since the spiked serum sample cannot be *a priori* considered commutable and appropriate for comparison with the control sample, the property of spiked serum sample, to be a substitute for a native serum with a high concentration of spiked analytes, was checked in the third survey. Using the same commutability criteria, the native serum from the third survey and the same spiked serum was evaluated for commutability. Only MPs showing commutability with native serum sample were further used in the second survey for evaluation.

All statistical analysis was performed using S-plus 8.0 (TIBCO Software Inc. Palo Alto, CA, USA) for Linux.

3.4 Analysis of statistically significant differences between native serum sample and lyophilized control samples

Analysis of statistically significant differences between native serum samples and lyophilized control samples was performed using analysis of variance (ANOVA).

Harmonisation assessment was performed by a one-way ANOVA with MP as a factor and using only the data of the serum sample and a correction for simultaneous hypothesis testing according to Tukey (109). Significant differences between MPs would indicate a lack of harmonisation. Assessing commutability of a control sample was performed by a two-way ANOVA with the laboratory as an extra random factor. Differences between the control and serum sample were compared between MPs. A correction for simultaneous hypothesis testing was applied according to Sidak (110). A significant difference between MPs of the differences between the two samples may indicate the lack of commutability of the control sample for those MPs. All statistically significant differences are calculated at the level of $P<0.05$.

3.5 False flagging method

To perform commutability evaluation based on pairwise comparison of MPs on serum and control sample, the *false flagging method* was introduced. Laboratories' results for each analyte are compared to the consensus target value of the MP group and APS of CROQALM (111) as presented in Table 4. The limits of CROQALM were chosen as the ones according to which the control samples would be evaluated since the same limits were used for individual results evaluation within the EQA scheme. APS of CROQALM are mostly based on biological variation data published by Ricos et al. (112), hosted and updated on Westgard webpage (60). For sodium and chloride, '*state of the art'* level is used, according to current technological possibilities.

Table 4. CROQALM analytical performance specifications

Analyte	**Acceptable deviation (%)**
GLUCOSE	7
BILIRUBIN	14
CREATININE	9
UREA	8
URATE	12
SODIUM	3
POTASSIUM	6
TOTAL CALCIUM	4
PHOSPHATE	10

Analyte	Acceptable deviation (%)
CHLORIDE	4
IRON	16
TOTAL CHOLESTEROL	9
HDL CHOLESTEROL	12
TRIGLYCERIDES	13
ALANINE AMINOTRANSFERASE	14
ASPARTATE AMINOTRANSFERASE	17
GAMMA GLUTAMYLTRANSFERASE	12
ALKALINE PHOSPHATASE	12
CREATINE KINASE	16
LACTATE DEHYDROGENASE	12
ALPHA AMYLASE	15
TOTAL PROTEINS	6

The results that exceed predefined limits are flagged and the flagging rate is calculated for each MP under evaluation.

A result is flagged when

$$\frac{|\text{laboratory result} - \text{consensus target value}|}{\text{consensus target value}} * 100 > \text{allowed deviation (\%)}$$

An EQA result that is obtained under optimal laboratory conditions should have only a small chance of being flagged. This probability is called the flagging rate and is given by:

$$P\left(X < consensus\ target\ value(1 - \frac{d}{100}) | X > consensus\ target\ value(1 + \frac{d}{100})\right) =$$

$$2 * P\left(X > consensus\ target\ value(1 + \frac{d}{100})\right) =$$

$$2 * P\left(\frac{X - consensus\ target\ value}{sd} > \frac{consensus\ target\ value * d}{sd * 100}\right) =$$

$$2 * P\left(Z > \frac{consensus\ target\ value * d}{sd * 100}\right)$$

where *X* stands for a reported EQA value, *d* stands for the value of the APS, *sd* stands for the standard deviation of the reported results and *Z* stands for a value of a standard normal distribution (with mean 0 and standard deviation 1), which is to be found in statistical textbooks or is given by appropriate statistical software.

The formula may also be rewritten as:

$$\text{flagging rate} = 2 * P\left(Z > \frac{d}{CV}\right)$$

with *CV* the coefficient of variation of the reported results. In other words, the larger d is with respect to the *CV* for a given MP, the lower the flagging rate.

For assessing commutability for a sample for two methods, two cases are considered: the first case is when the results of the two MPs are joined into one peer group and the second case is when the results of the two MPs are evaluated in two separate peer groups. The differences in flagging rate are calculated between the case where the MPs are joined in one group and the case when they are in separate groups. This calculation is performed for the control sample on the one hand and for the serum sample on the other hand. The control sample is considered as commutable if the differences obtained for the control sample and for the serum sample are close to each other. Flagging rate differences that exceed the maximum allowable rate of 20% for any pairwise comparison of MPs is considered a false flagging rate and set as a commutability limit of control materials (Figure 10). By allowing the 5% change on each side of the curve, the total change for one curve, or MP, is 10%, and for two MPs evaluated in each set of analysis, this yields 20% change in flagging rate.

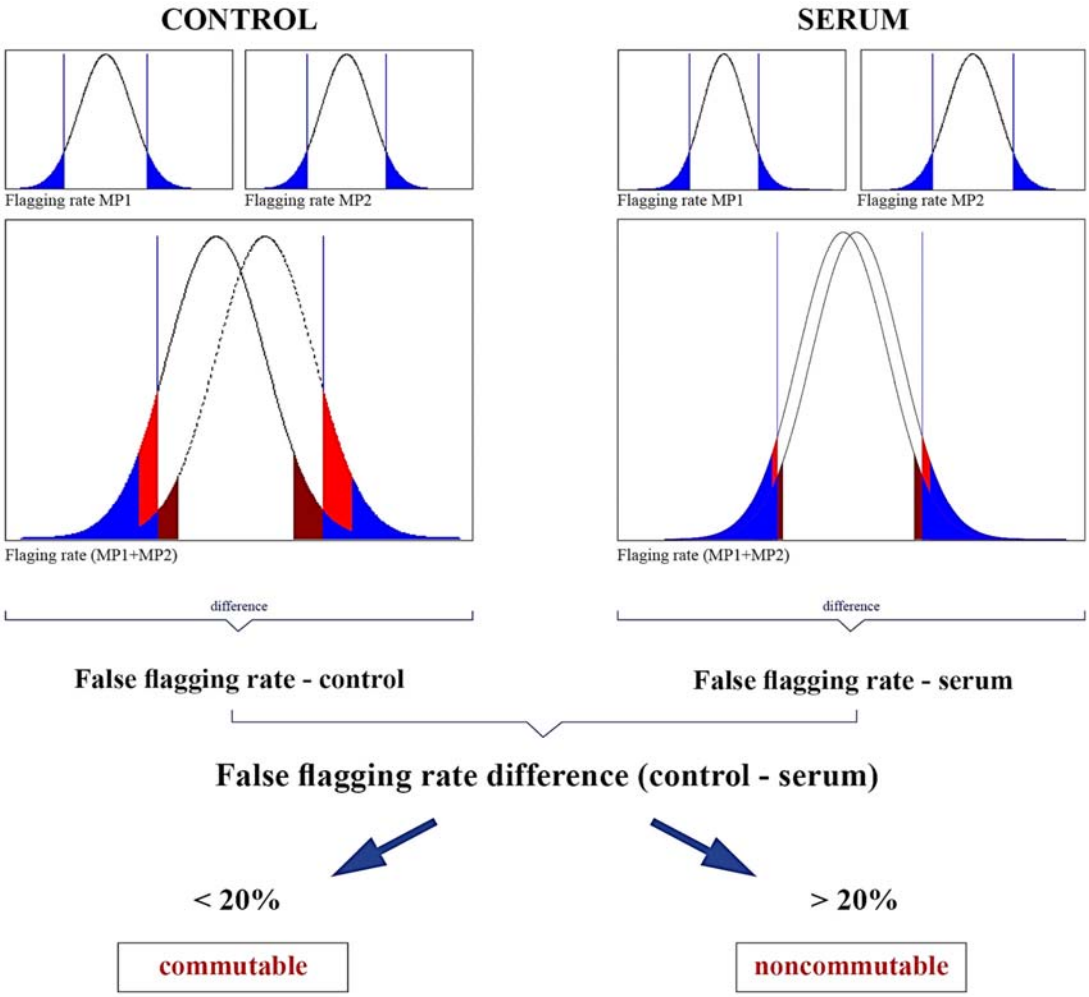


Figure 10. Scheme of **a** false flagging method for commutability evaluation**.**

The probability that the limit of 20% of the false flagging rate is exceeded can be calculated using the mean, standard deviation and number of data in each group. Considering the fact that the mean and the standard deviation of the certain data series are variable and slightly different each time the data is collected, it is important to calculate the chance that the upper limit of 20% points false flagging rate would ever be exceeded, taking into account the current mean, standard deviation and the number of data. This probability is obtained using bootstrapping (113). Starting from the certain data series, a new series is made that consists of random selecting (sampling) data that are the part of initial data series, in which a certain value can occur more than once. The probability of changed flagging is calculated using 1000 bootstrapped samples of the originally reported results.

The commutability evaluation of lyophilised control samples for each pairwise comparison of MPs is performed using the following approach:

1. **Serum sample:** For each pair of MPs calculate the consensus target value and standard deviation two times: once for each MP apart by using the consensus target value for each MP apart, and once for the two MPs together using a consensus target value calculated after joining the results of the two MPs into one group. Calculate in both cases the probability of flagging according to the consensus target value, the standard deviation, and defined APS. False flagging is defined as the difference in flagging rate between the case when the MPs are evaluated apart and when they are put into one group.
2. **Lyophilised commercial control sample:** For each pair of MPs calculate the consensus target value and standard deviation two times: once for each MP apart by using the consensus target value for each MP apart, and once for the two MPs together using a consensus target value calculated after joining the results of the two MPs into one group. Calculate in both cases the probability of flagging according to the consensus target value the standard deviation, and defined APS.

   The false flagging rate between two MPs observed on lyophilised control samples should be similar to the false flagging rate observed on serum sample if the control material is commutable. The maximum allowed difference in the false flagging rate of control material compared to the serum sample was set to 20% points.
3. Create 1000 bootstrapped samples (set of results) for each MP and EQA sample (serum and lyophilised commercial control sample) - sampling with replacement. Repeat steps 1 and 2 for each bootstrapped sample. Calculate the false flagging rate for each bootstrapped

sample as the difference in flagging rate between lyophilised control samples and serum samples.

4. Calculate the percentage commutability as the percentage of bootstrapped samples not exceeding the predefined limit of 20%point difference falsely flagged results between control and fresh sample.

Lyophilised commercial control samples are defined as commutable for assessed MPs combination if percentage commutability is ≥ 95%. The control samples are defined as noncommutable if the percentage commutability is < 95%. The 95%-acceptance criteria were chosen as the usual 95%-significance confidence level used in statistical inference.

To quantify the initial harmonisation between two MPs, the same logic of falsely flagged results is applied. If the results from two MPs are harmonised, the false flagging rate does not change substantially if the methods are joined into one group and individual results evaluated according to unique target value compared to a separate evaluation per MP. The change in false flagging rate above the predefined limit can be observed for nonharmonised MPs, yielding a larger proportion of laboratories to be flagged when two groups are joined. The initial harmonisation between MPs is evaluated using the analysis results of a serum sample. The change in flagging rate when the results from two MPs are joined and evaluated within one peer group is considered as false flagging rate. The limit of 20% of the false flagging rate is used for defining harmonisation between methods. If the results from two MPs differ substantially such that joining groups results in more than 20% falsely flagged results, the MPs are considered nonharmonised. The flagging rate within the predefined limit of 20% is observed for harmonised MPs. The percentage harmonisation is calculated as the percentage of bootstrapped samples not exceeding the false flagging limit on native serum samples. MPs are defined as harmonised if percentage harmonisation is ≥ 95%.

Based on the percentage of MP combinations commutable for an analyte, the analyte-related commutability of lyophilised control materials are further classified as follows: (1) *Full commutability*, commutable for all MPs combinations used in measurements of corresponding analyte; (2) *High commutability*, noncommutable for < 20% MPs combinations used in measurement of stated analyte; (3) *Moderate commutability*, noncommutable for 20 - 60% MPs combinations used in measurement of stated analyte and (4) *Noncommutability* (NC), noncommutable for > 60 % MPs combinations used in the measurement of stated analyte. The criteria for this classification was subjectively chosen in order to allocate the control samples to different classes according to the need for future evaluations, where fully commutable and

noncommutable controls would not need any future assessment of commutability. In the attempt to be able to compare commutability results of evaluated MP pairs using regression analysis and proposed false flagging method for commutability evaluation, the classes high commutability and moderate commutability were introduced.

4. RESULTS

4.1 Commutability evaluation of control samples using regression analysis

As the first step in commutability evaluation of control samples, regression analysis is performed. Twenty to twenty-two residual patient serum samples spanning the broad concentration range of evaluated analytes were measured in the same run on the instrument as three control samples, according to the procedure 3.2 in Material and methods section. The results for sodium, potassium and chloride for Cobas Integra were excluded from further analysis due to large unexplained differences observed for control samples. The lyophilised control sample from EQA survey 2 was excluded from the analysis of commutability for triglycerides because the results for that control sample largely exceeds the concentration range measured in patient samples.

Median concentrations of control samples measured in participating laboratories are listed in Table 5. The concentration ranges for two controls (C1/2016 and C3/2016) represent the normal or low pathological concentration ranges considering appropriate reference intervals for stated analytes, whereas the concentration ranges in the control C2/2016 correspond to the pathological concentration levels.

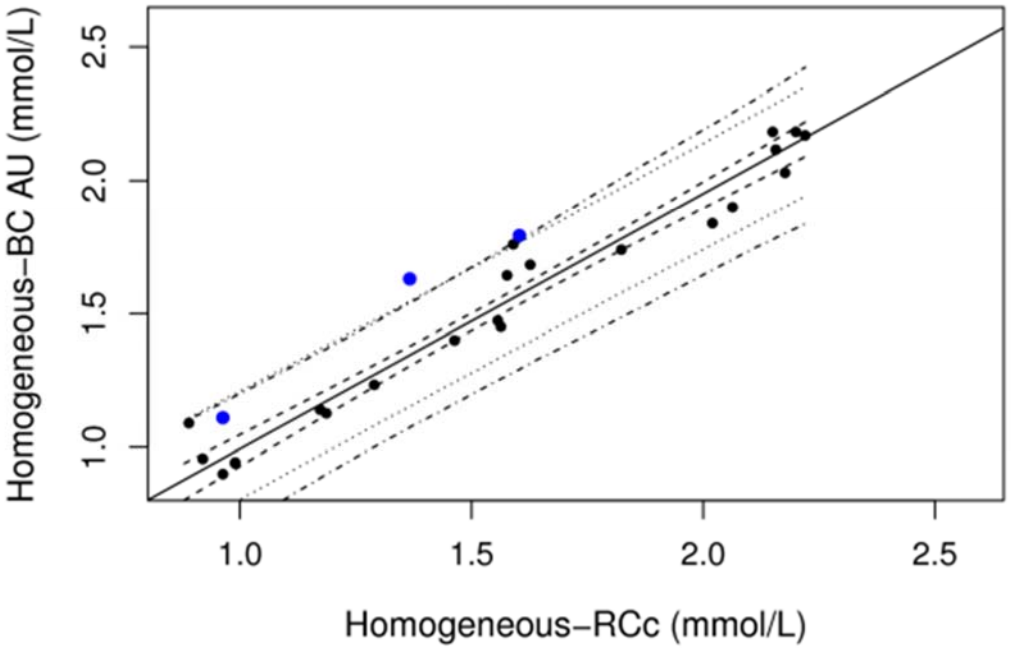
When the measurement results of patient samples were plotted on the appropriate graphs showing regression line and 95% prediction interval, we observed too many results were outside of the proposed interval. A number of patients outside the limits of prediction interval was 419/2280 (18.4%) for all pairwise combinations of MPs. This clearly showed that the width of the prediction interval suggested in CLSI guideline was too narrow, not consisted of neither nearly 95% of measured results from patients. The percentage of patient results outside of the prediction interval of the Deming regression line was as high as 52.9% for some pairwise combinations of MPs. The range of total patients outside the limits of prediction interval for each analyte and all evaluated MP pairs was 10.9 - 39.2%. Such observation led to the conclusion that the regression analysis used for evaluation of commutability of control samples has to be changed since it is the prediction interval itself that is used as a commutability criterion.

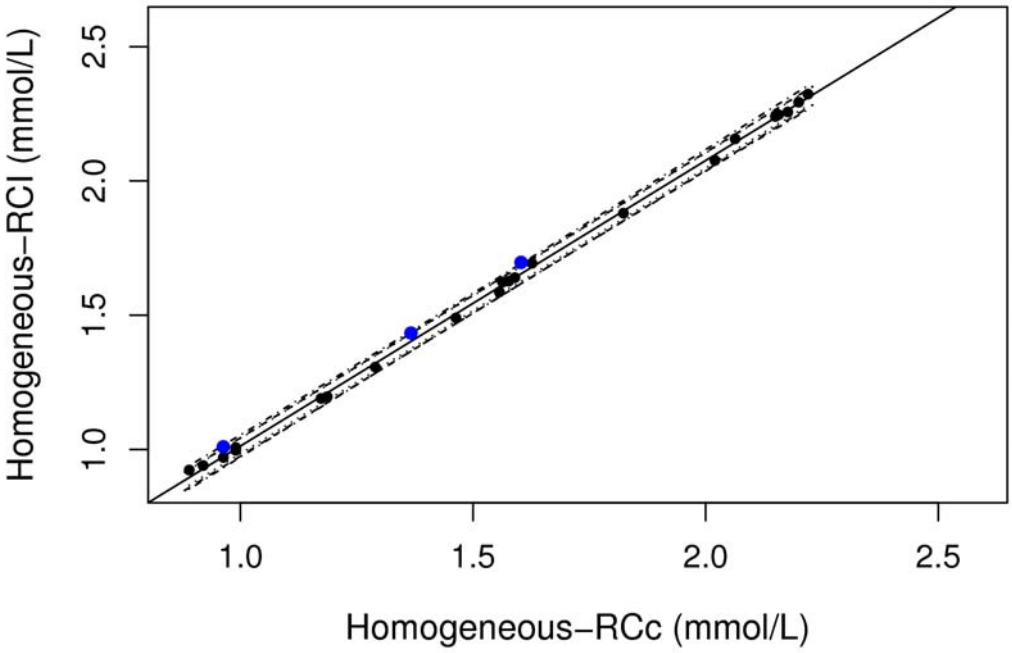
Table 5. Median concentrations of analytes in control samples assessed in CLSI commutability study

Analyte (Units)	Method	C1/2016	C2/2016	C3/2016
Alanine aminotransferase (U/L)	IFCC Photometry UV	41	140	30
Aspartate aminotransferase (U/L)	IFCC Photometry UV	42	216	38
Chloride (mmol/L)	Indirect ISE	112	132	103
Cholesterol (mmol/L)	CHOD-PAP	4.1	6.0	6.5
Creatinine (µmol/L)	Compensated Jaffe	78	248	185
Gamma glutamyltransferase (U/L)	IFCC	36	144	59
Glucose (mmol/L)	Hexokinase	4.1	10.2	4.4
HDL cholesterol (mmol/L)	Homogenous enzymatic	1.1	1.6	1.8
Potassium (mmol/L)	Indirect ISE	3.5	5.8	3.9
Sodium (mmol/L)	Indirect ISE	138	166	145
Triglycerides (mmol/L)	GPO-PAP	1.1	4.6	2.1
Urea (mmol/L)	Urease, GLDH	4.9	13.2	5.6

C1/2016, C2/2016, C3/2016 – commercial control samples evaluated in three EQA surveys, IFCC – International Federation of Clinical Chemistry and Laboratory Medicine, ISE – Ion Selective Electrode, CHOD-PAP – Cholesterol oxidase/peroxidase- phenol/4-aminophenazone, GPO-PAP - Glycerol phosphate oxidase/peroxidase - phenol/4-aminophenazone, GLDH – Glutamate dehydrogenase.

The choice of regression analysis to be used for analysis was made after graphical inspection of the width of prediction intervals proposed for Passing and Bablok and simple linear regression analysis (Figure 11) and by calculating the number of patient results that would fit into the proposed interval. The percentage of results being outside the limits of 95% prediction interval was 1.3 using Passing and Bablok regression analysis and 2.9 using ordinary linear regression. For these reasons, the linear regression was further used in commutability evaluation of control samples.


Homogeneous−BC AU (mmol/L)
1.0
1.5
2.0
2.5
Homogeneous−RCc (mmol/L)


Homogeneous−RCI (mmol/L)
1.0
1.5
2.0
2.5
Homogeneous−RCc (mmol/L)

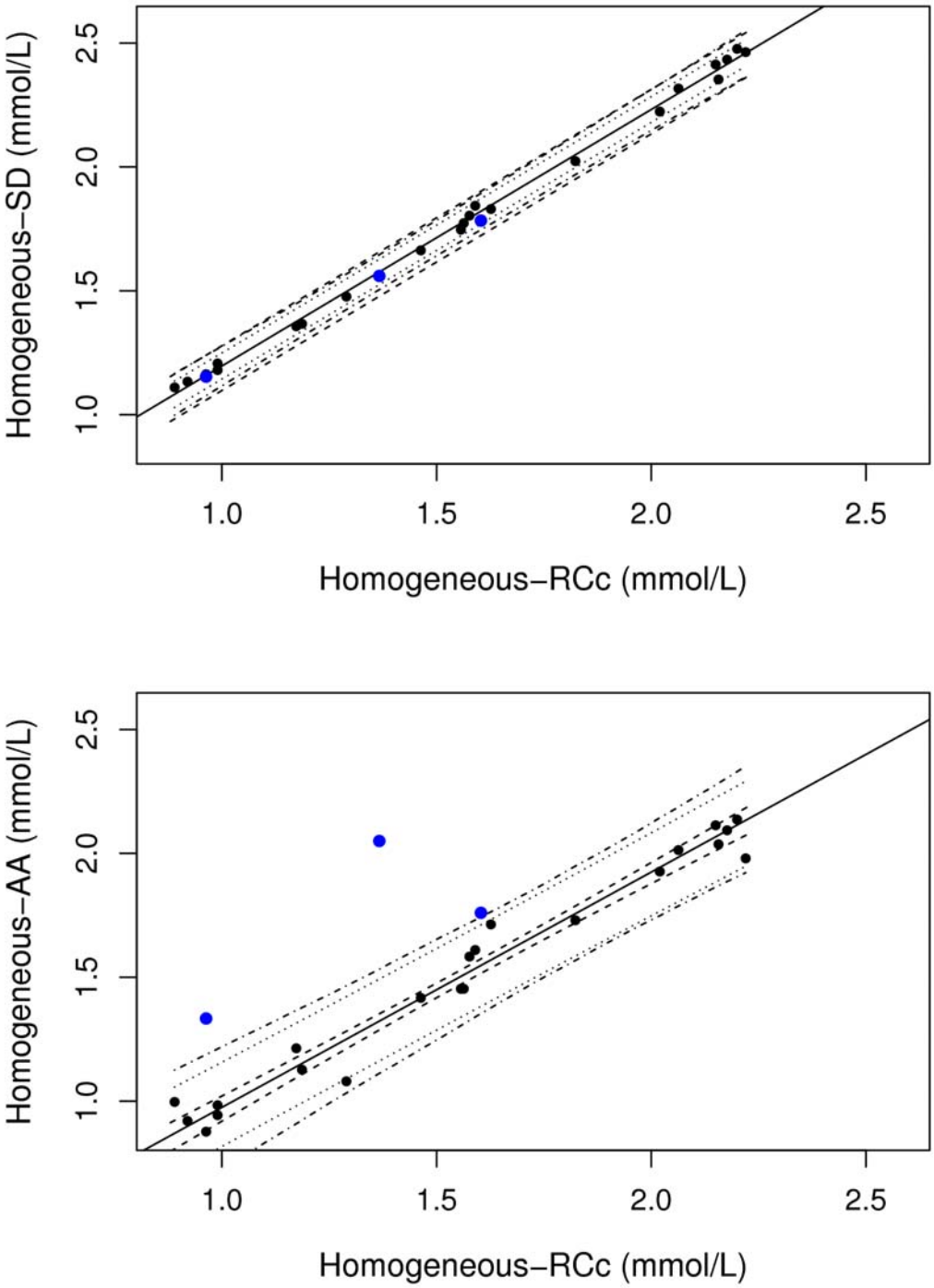
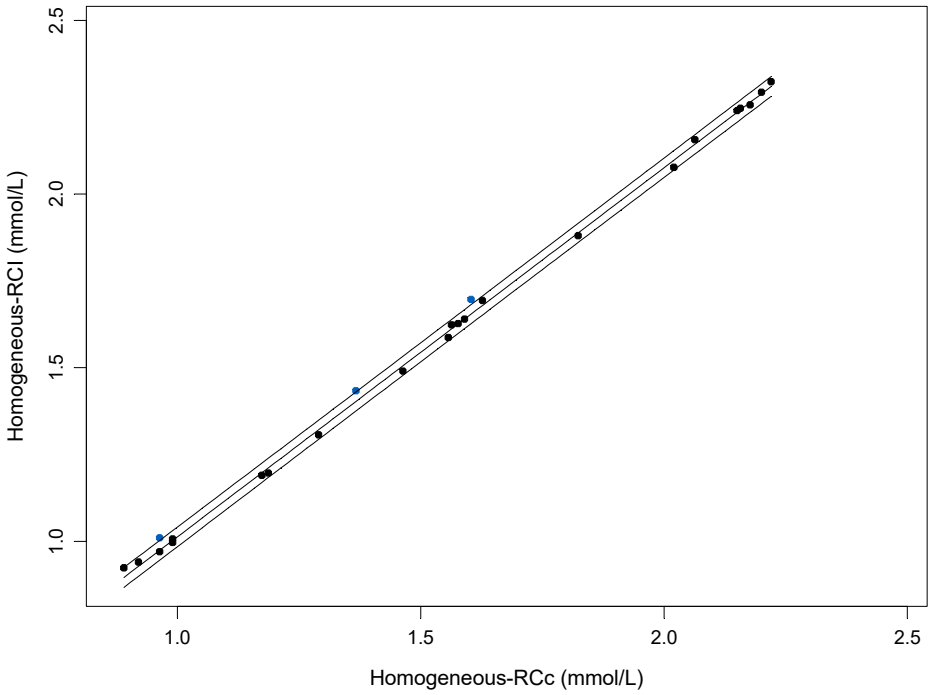
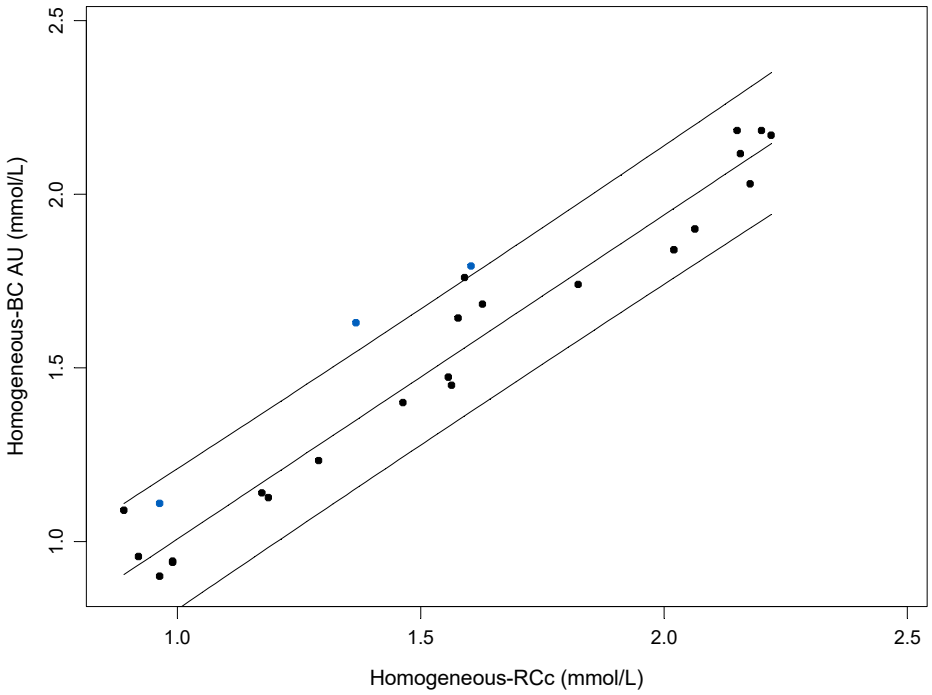
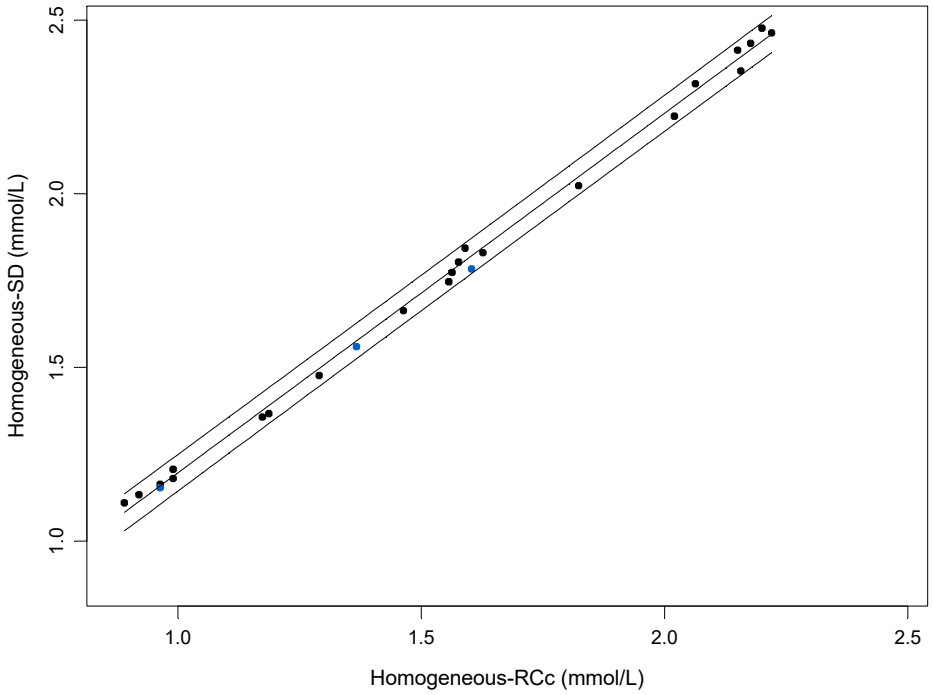


Figure 11. Commutability evaluation of control samples for HDL cholesterol using regression analysis: different kinds of 95% prediction intervals. The blue dots present the measurement results of control samples (C1-C3/2016) using homogenous enzymatic method with all evaluated instruments compared to Roche Cobas c (RCc): Beckman Coulter AU (BC AU), Roche Cobas Integra (RCI), Siemens Dimension (SD) and Abbott Architect (AA). The graphs show the Deming regression line (black solid line) and the 95% prediction intervals recommended in CLSI EP14-A3 (Deming regression - dashed lines), CLSI EP14-A2 (simple linear regression - dotted lines) and Passing and Bablok regression (dot-dashed lines).

An illustrative example of commutability evaluation recommended by CSLI is presented on the example of HDL cholesterol in Figure 12. The results of measurement of HDL cholesterol in patient sera (black dots) on different MPs were plotted on the corresponding graphs for each MPs pair and the regression line, together with the 95% prediction interval. The width of the prediction interval depends on the uncertainty around the relationship between the measurements of the serum samples by two MPs. The blue dots representing the measurement results of control samples using the same MPs are plotted on the same graphs. The control samples that fall outside of the 95% prediction interval calculated in regression analysis are considered noncommutable. Commutability results for corresponding MPs combinations are presented in Table 6. All three control samples are noncommutable for homogeneous enzymatic method on instruments Roche Cobas c and Roche Cobas Integra, as well as a homogeneous enzymatic method on instruments Roche Cobas c and Abbott Architect. On the contrary, the controls are fully commutable for homogeneous enzymatic method on instruments Roche Cobas c and Siemens Dimension.




Homogeneous-BC AU (mmol/L)
Homogeneous-RCc (mmol/L)
1.0
1.5
2.0
2.5


Homogeneous-SD (mmol/L)
Homogeneous-RCc (mmol/L)
1.0
1.5
2.0
2.5

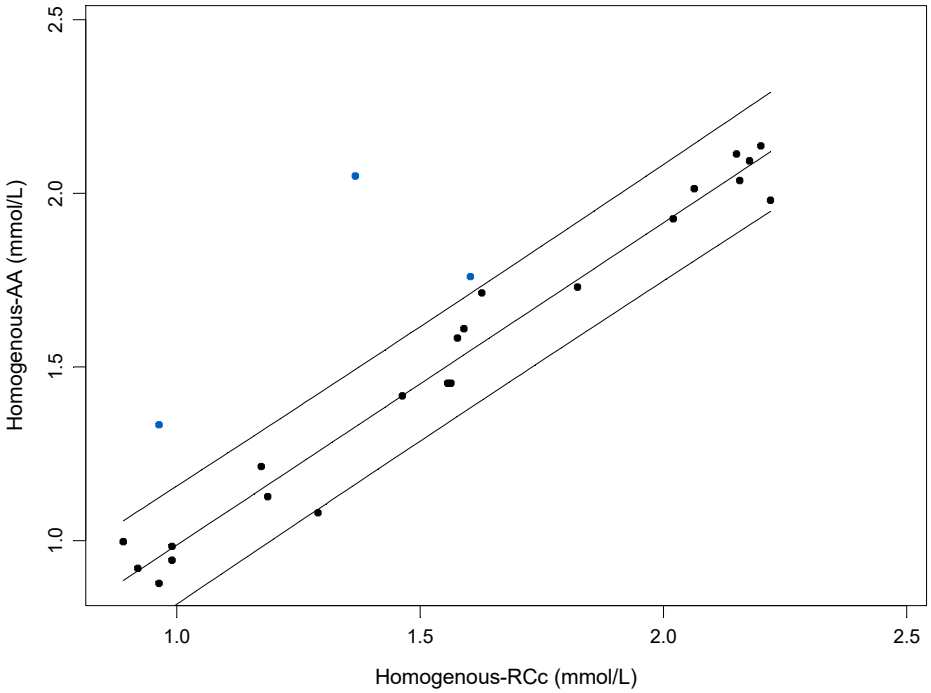


Figure 12. Commutability assessment of control samples for HDL cholesterol measurement using linear regression analysis. The graphs show the regression line (black solid line) and the 95% prediction interval around the regression line (black dashed lines) of measurement results of patient serum samples (black circles). The blue dots present the measurement results of control materials (C1-C3/2016) using homogenous enzymatic method with all assessed instruments compared to Roche Cobas c (RCc): Roche Cobas Integra (RCI), Beckman Coulter AU (BC AU), Siemens Dimension (SD) and Abbott Architect (AA).

Table 6. Commutability results of regression analysis of control materials for HDL cholesterol using homogeneous enzymatic assays for all instruments compared to Roche Cobas c.

Control sample		Roche Cobas Integra	Abbott Architect	Beckman Coulter AU	Siemens Dimension
C1/2016	**Roche Cobas c**	NC	NC	C	C
C2/2016		NC	NC	NC	C
C3/2016		NC	NC	NC	C

NC – noncommutable, C – commutable

The results of commutability assessment of EQA control samples (C1/2016, C2/2016 and C3/2016) for investigated analytes using routine MPs on five instruments are presented in Tables 7 - 9. All three control samples were found fully commutable for all combinations of MPs used for measurement of potassium, sodium and GGT. Considering the number of commutable decisions throughout all MPs accessed, the controls are also highly commutable for AST (only one *noncommutable* decision) and triglycerides in the controls of the normal range.

The control C1/2016 showed to be noncommutable for 5 MP combinations for HDL cholesterol, total cholesterol and for 6 MP combination for glucose (Table 7). Out of the total 7 MPs combinations assessed for chloride, the control sample was noncommutable for 3 MPs combinations. Except for one MPs combination for ALT, C1/2016 showed to be commutable for all MP used for measuring potassium, sodium, creatinine, urea, GGT, AST and ALT.

The control C2/2016 showed similar patterns of noncommutability as C1/2016 for HDL-cholesterol, total cholesterol and glucose, with noncommutability being found for even more MPs combinations. The commutability of the control sample was somewhat better for chloride, but markedly worse for creatinine, with 7 noncommutable MPs combinations of total 10.

Although being from the different manufacturer, the control 3/2106 also showed high noncommutability for HDL cholesterol, total cholesterol, glucose, chloride and creatinine, with the addition to being also noncommutable for 6 MPs combinations for ALT.

Table 7. Summary of commutability conclusions according to regression analysis for EQA control sample C1/2016

Instrument pair	HDL	Triglycerides	Cholesterol	Glucose	Chloride	Potassium	Sodium	Creatinine	Urea	GGT	AST	ALT
AA - BC	NC	C	C	NC	C	C	C	C	C	C	C	C
AA - RCc	NC	C	C	C	C	C	C	C	C	C	C	C
AA - RCI	NC	C	NC	C	/	/	/	C	C	C	C	C
AA - SD	NC	C	NC	NC	NC	C	C	C	C	C	C	C
BC - RCc	C	C	C	NC	NC	C	C	C	C	C	C	C
BC - RCI	C	NC	C	NC	/	/	/	C	C	C	C	NC
BC - SD	C	C	NC	C	NC	C	C	C	C	C	C	C
RCc- RCI	NC	C	C	C	/	/	/	C	C	C	C	C
RCc - SD	C	C	NC	NC	C	C	C	C	C	C	C	C
RCI - SD	C	NC	NC	NC	/	/	/	C	C	C	C	C

NC – noncommutable, C – commutable, AA - Abbott Architect c4000, BC - Beckman Coulter AU, RCc - Roche Cobas 6000 c501, RCI - Roche Cobas Integra 400, SD - Siemens Dimension Xpand

Table 8. Summary of commutability conclusions according to regression analysis for EQA control sample C2/2016

Instrument pair	HDL	Cholesterol	Glucose	Chloride	Potassium	Sodium	Creatinine	Urea	GGT	AST	ALT
AA - BC	NC	C	C	C	C	C	NC	C	C	C	C
AA - RCc	NC	NC	NC	C	C	C	NC	NC	C	C	C
AA - RCI	NC	C	C	/	/	/	NC	NC	C	C	C
AA - SD	NC	NC	NC	NC	C	C	NC	C	C	C	C
BC - RCc	NC	NC	NC	C	C	C	NC	C	C	C	C
BC - RCI	NC	C	C	/	/	/	NC	C	C	C	C
BC - SD	NC	NC	NC	C	C	C	C	C	C	C	C
RCc- RCI	NC	C	C	/	/	/	C	C	C	NC	C
RCc - SD	C	NC	NC	C	C	C	NC	C	C	C	C
RCI - SD	NC	NC	NC	/	/	/	C	C	C	C	C

NC – noncommutable, C – commutable, AA - Abbott Architect c4000, BC - Beckman Coulter AU, RCc - Roche Cobas 6000 c501, RCI - Roche Cobas Integra 400, SD – Siemens Dimension Xpand

Table 9. Summary of commutability conclusions according to regression analysis for EQA control sample C3/2016

Instrument pair	HDL	Triglycerides	Cholesterol	Glucose	Chloride	Potassium	Sodium	Creatinine	Urea	GGT	AST	ALT
AA - BC	C	C	NC	C	C	C	C	NC	C	C	C	C
AA - RCc	NC	C	C	NC	NC	C	C	NC	C	C	C	C
AA - RCI	NC	C	C	C	/	/	/	NC	C	C	C	C
AA - SD	NC	C	NC	NC	NC	C	C	C	C	C	C	NC
BC - RCc	NC	C	NC	NC	NC	C	C	C	C	C	C	NC
BC - RCI	NC	NC	NC	C	/	/	/	C	C	C	C	C
BC - SD	NC	C	C	C	NC	C	C	NC	C	C	C	NC
RCc- RCI	NC	C	C	C	/	/	/	C	C	C	C	NC
RCc - SD	C	C	NC	NC	NC	C	C	NC	C	C	C	NC
RCI - SD	NC	NC	NC	NC	/	/	/	NC	C	C	C	NC

NC – noncommutable, C – commutable, AA - Abbott Architect c4000, BC - Beckman Coulter AU, RCc - Roche Cobas 6000 c501, RCI - Roche Cobas Integra 400, SD - Siemens Dimension Xpand

4.2 Commutability evaluation of control samples within EQA

4.2.1 Statistical significance of differences between control and human samples

Three serum and control samples were analysed by 180 – 184 laboratories participating in the EQA surveys at the same time using appropriate MPs. The measurement results for each analyte on these two types of samples were recorded and MPs with at least 6 results were further assessed in commutability evaluation. The total of 143 MPs used in the measurement of 22 analytes from the biochemistry module in CROQALM were formed.

The difference between measurement results for each analyte with control and serum samples were calculated. It was expected that for commutable control samples, these differences would be similar, not significantly different. In order to evaluate these differences among various MPs used for measurement of an analyte, the ANOVA approach described in section 3.4 was used to evaluate observed differences between differences in the measurement of the analyte with control and serum samples.

Figures 13-15 represent the differences of medians obtained for control and serum samples in dependence on MP. It can be seen from the figures that parallel lines, corresponding to measurements with different MPs, are expected with commutable control samples. Significant differences among measurements of serum and control samples result in lines getting closer or further apart from each other. The differences can also be presented by means of vertical bars, where the height of the bars represents the difference between the results of serum and control sample. The graphs are also showing the degree of harmonisation between MPs, presented as differences between results from MPs on serum samples.

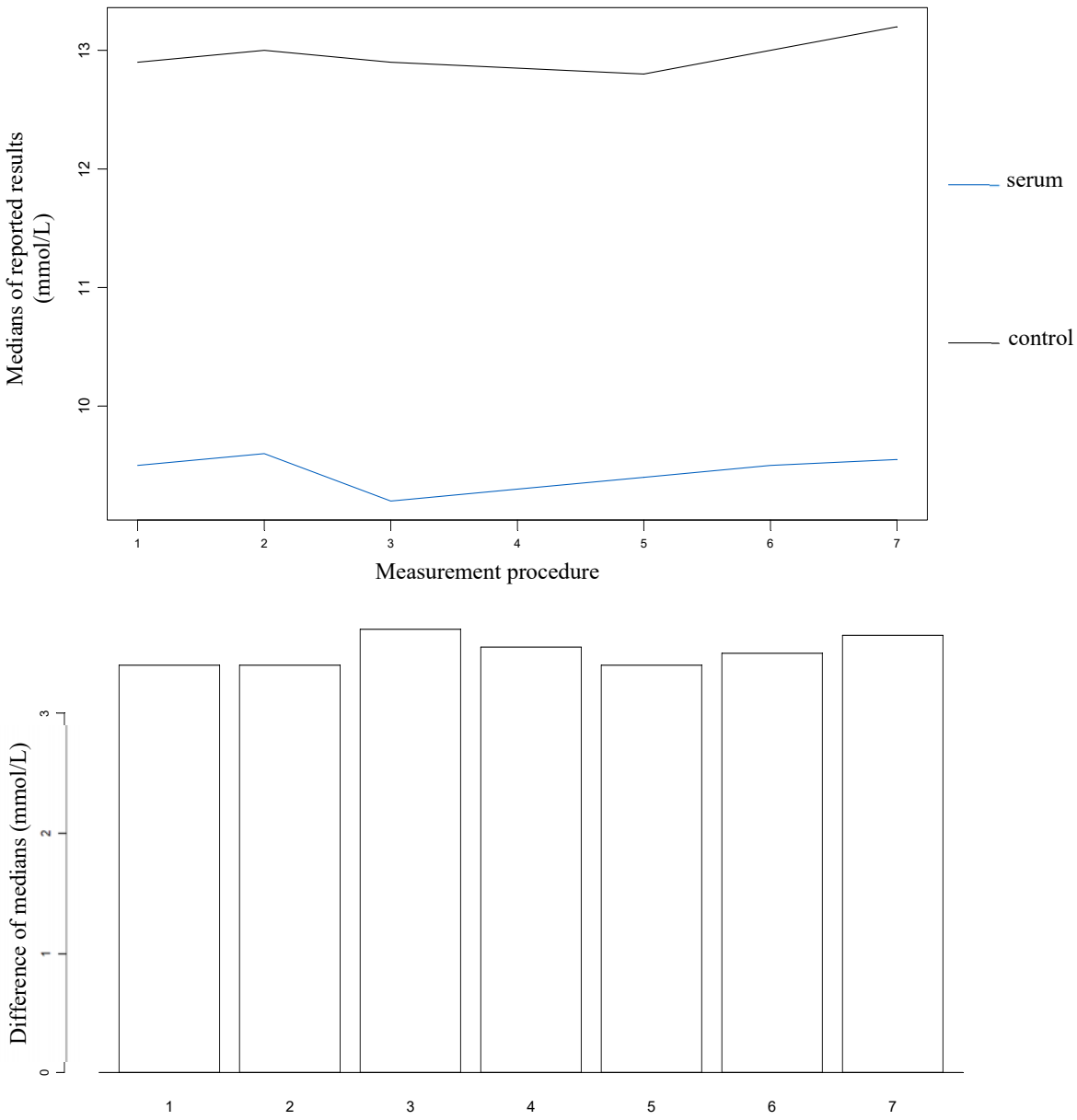


Figure 13. Graphical presentation of differences among measurements of control and serum samples for urea. Codes 1-7 represent assessed MPs: Urease, GLDH -AA (1); Urease, GLDH -BC AU (2); Urease, GLDH -HP (3); Urease, GLDH -RCc (4); Urease, GLDH -RCI (5); Urease, GLDH -RH (6); Urease, GLDH -SD (7). The upper graph shows the linear plot of medians of reported results for each MP, whereas the differences of these medians are presented as bar plots on the graph below.

Figure 13 gives a strong indication of commutability of control sample C2/2016 for all MPs used for urea measurement. This observation is also confirmed by the ANOVA result of differences between C2/2016 and serum differences on each MP, as presented in Table 10.

Table 10. The statistical analysis of differences between C2/2016 control and serum sample differences for each pairwise comparison of MPs used for measurement of urea.

MPs for comparison	Diff*	95% CI	P-value	Commutability
Urease,GLDH –AA/Urease,GLDH -BC AU	0.06	(-0.21 - 0.32)	0.9953	C
Urease,GLDH –AA/Urease,GLDH -HP	0.01	(-0.41 - 0.42)	0.9999	C
Urease,GLDH –AA/Urease,GLDH -RCc	-0.03	(-0.36 - 0.29)	0.9999	C
Urease,GLDH –AA/Urease,GLDH -RCI	0.04	(-0.29 - 0.36)	0.9999	C
Urease,GLDH –AA/Urease,GLDH -RH	0.06	(-0.24 - 0.36)	0.9975	C
Urease,GLDH –AA/Urease,GLDH -SD	-0.01	(-0.36 - 0.34)	0.9999	C
Urease,GLDH -BC AU/Urease,GLDH -RCc	-0.09	(-0.33 - 0.15)	0.917	C
Urease,GLDH -BC AU/Urease,GLDH -RCI	-0.02	(-0.25 - 0.21)	0.9999	C
Urease,GLDH -BC AU/Urease,GLDH -RH	0	(-0.2 - 0.2)	0.9999	C
Urease,GLDH -BC AU/Urease,GLDH -SD	-0.07	(-0.33 - 0.2)	0.9886	C
Urease,GLDH –HP/Urease,GLDH -RCc	-0.04	(-0.44 - 0.36)	0.9999	C
Urease,GLDH –HP/Urease,GLDH -RCI	0.03	(-0.36 - 0.43)	0.9999	C
Urease,GLDH -RCc /Urease,GLDH -RCI	0.07	(-0.23 - 0.38)	0.9921	C
Urease,GLDH –RCc/Urease,GLDH -RH	0.09	(-0.19 - 0.37)	0.9587	C
Urease,GLDH –RCc/Urease,GLDH -SD	0.02	(-0.3 - 0.35)	0.9999	C
Urease,GLDH –RCI/Urease,GLDH -RH	0.02	(-0.26 - 0.3)	0.9999	C
Urease,GLDH –RCI/Urease,GLDH -SD	-0.05	(-0.37 - 0.27)	0.9995	C
Urease,GLDH –RH/Urease,GLDH -SD	-0.07	(-0.37 - 0.23)	0.9939	C

Diff. - difference between control and serum sample differences for each MPs pair, C – commutable, NC - noncommutable

A larger difference between cholesterol measurements for control and serum sample can be observed in Figure 14 for CHOD-PAP-SD compared to other MPs. It can be seen that the difference for that instrument is substantially larger than for the other instruments used for cholesterol measurement.

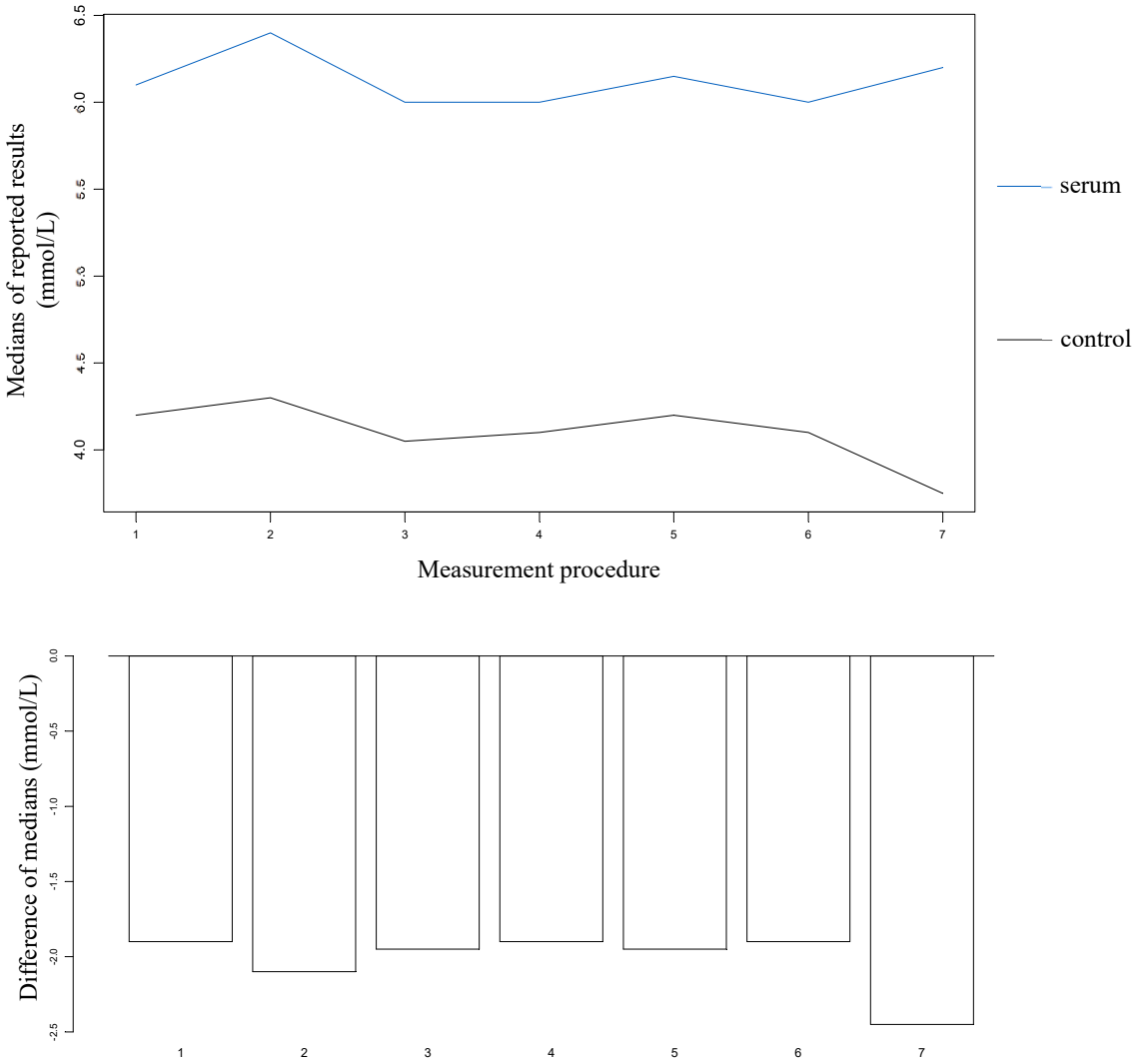


Figure 14. Graphical presentation of differences among measurements of control and serum sample for cholesterol. Codes 1-7 represent assessed MPs: CHOD-PAP- AA (1), CHOD-PAP- BC AU (2), CHOD-PAP-HP (3), CHOD-PAP - RCI (4), CHOD-PAP - RCc (5), CHOD-PAP -RH (6), CHOD-PAP - SD (7). The upper graph shows the linear plot of medians of reported results for each MP, whereas the differences of these medians are presented as bar plots on the graph below.

The statistically significant difference of differences between control and serum sample was observed for all pairwise comparison of MPs including CHOD-PAP- SD, leading to the conclusion of noncommutability of this MP for the assessed control sample (Table 11). Table 11 also reveals that the MP comparisons including CHOD-PAP-BC AU show noncommutability in 4/6 pairwise comparisons with this instrument. Overall, the control sample is noncommutable for 9/21 MP pairs evaluated.

Table 11. The statistical analysis of differences between C1/2016 control and serum sample differences for each pairwise comparison of MPs used for measurement of cholesterol.

MPs for comparison	Diff*	95% CI	P-value	Commutability
CHOD-PAP - AA- CHOD-PAP - BC AU	0.12	(-0.01-0.25)	0.0734	C
CHOD-PAP - AA- CHOD-PAP - HP	-0.04	(-0.24-0.16)	0.9965	C
CHOD-PAP - AA- CHOD-PAP - RCc	-0.03	(-0.19-0.13)	0.9982	C
CHOD-PAP - AA- CHOD-PAP - RCI	-0.04	(-0.19-0.12)	0.9925	C
CHOD-PAP - AA- CHOD-PAP - RH	-0.05	(-0.19-0.1)	0.9659	C
CHOD-PAP - AA- CHOD-PAP - SD	0.46	(0.29-0.63)	0.0001	NC
CHOD-PAP - BC AU- CHOD-PAP - HP	-0.16	(-0.33-0.01)	0.0626	C
CHOD-PAP - BC AU- CHOD-PAP - RCc	-0.15	(-0.27-(-0.03))	0.0033	NC
CHOD-PAP - BC AU- CHOD-PAP - RCI	-0.15	(-0.26-(-0.05))	0.0002	NC
CHOD-PAP - BC AU- CHOD-PAP - RH	-0.16	(-0.26-(-0.07))	0.0001	NC
CHOD-PAP - BC AU- CHOD-PAP - SD	0.34	(0.21-0.47)	0.0001	NC
CHOD-PAP - HP- CHOD-PAP - RCc	0.01	(-0.18-0.21)	0.9999	C
CHOD-PAP - HP- CHOD-PAP - RCI	0.01	(-0.18-0.19)	0.9999	C
CHOD-PAP - HP- CHOD-PAP - RH	0	(-0.18-0.18)	0.9999	C
CHOD-PAP - HP- CHOD-PAP - SD	0.5	(0.3-0.7)	0.0001	NC
CHOD-PAP - RCc- CHOD-PAP - RCI	-0.01	(-0.15-0.14)	0.9999	C
CHOD-PAP - RCc- CHOD-PAP - RH	-0.02	(-0.15-0.12)	0.9999	C
CHOD-PAP - RCc- CHOD-PAP - SD	0.49	(0.32-0.65)	0.0001	NC
CHOD-PAP - RCI- CHOD-PAP - RH	-0.01	(-0.13-0.11)	0.9999	C
CHOD-PAP - RCI- CHOD-PAP - SD	0.49	(0.34-0.65)	0.0001	NC
CHOD-PAP - RH- CHOD-PAP - SD	0.5	(0.36-0.65)	0.0001	NC

Diff. - difference between control and serum sample differences for each MP pair, C – commutable, NC - noncommutable

The differences among measurements of control and serum sample for creatinine are presented in Figure 15. Besides indicating noncommutability of C1/2016 for some MPs combinations, graphical presentation of results strongly suggests the lack of harmonisation of MPs used for creatinine measurement considering the differences of results seen on the serum samples (blue line). As presented in Table 12, the control C1/2016 was found noncommutable for 6/36 MPs evaluated.

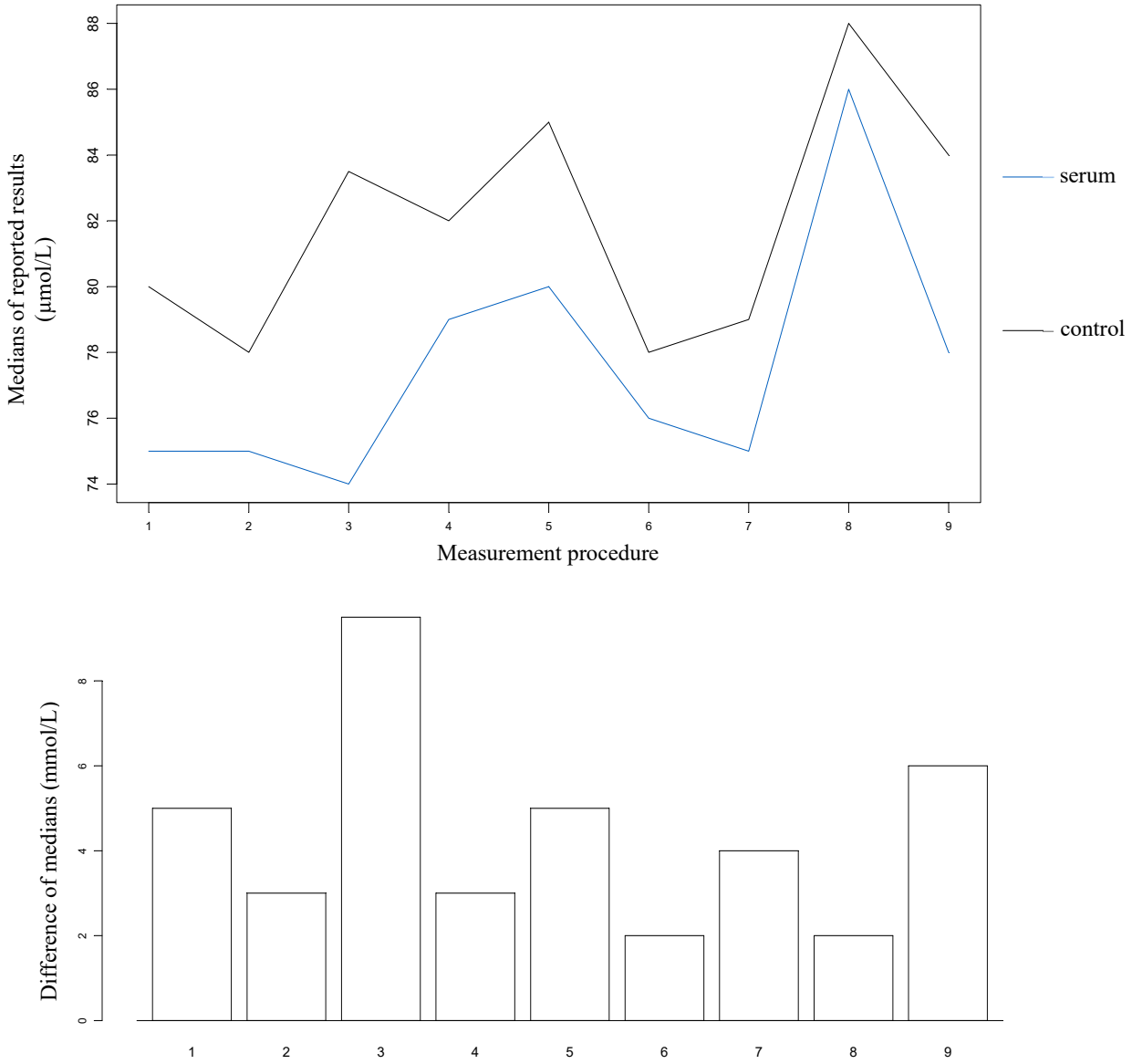


Figure 15. Graphical presentation of differences among measurements of control and serum sample for creatinine. Codes 1-7 represent assessed MPs: Compensated Jaffe- AA (1), Compensated Jaffe- BC AU (2), Compensated Jaffe- RCc (3), Compensated Jaffe- RCI (4), Compensated Jaffe- RH (5), Compensated Jaffe- SD (6), Enzymatic method-BC AU (7), Non-compensated Jaffe- BC AU (8), Non-compensated Jaffe- RH (9). The upper graph shows the linear plot of medians of reported results for each MP, whereas the differences of these medians are presented as bar plots on the graph below.

Table 12. The statistical analysis of differences between C1/2016 control and serum sample differences for each pairwise comparison of MPs used for measurement of creatinine.

MPs for comparison	Diff.	95% CI	P-value	Comm.
Compensated Jaffe-AA- Compensated Jaffe-BC AU	2.4	(-2.4-7.2)	0.8152	C
Compensated Jaffe-AA- Compensated Jaffe-RCc	-5.3	(-11.01-0.41)	0.0821	C
Compensated Jaffe-AA- Compensated Jaffe-RCI	-0.58	(-6.11-4.94)	0.9999	C
Compensated Jaffe-AA- Compensated Jaffe-RH	-1.09	(-6.42-4.24)	0.9994	C
Compensated Jaffe-AA- Compensated Jaffe-SD	3.22	(-2.6-9.05)	0.7167	C
Compensated Jaffe-AA- Enzymatic method-BC AU	0.09	(-5.6-5.77)	0.9999	C
Compensated Jaffe-AA- Non-compensated Jaffe-BC AU	1.66	(-3.45-6.76)	0.9835	C
Compensated Jaffe-AA- Non-compensated Jaffe-RH	-0.75	(-6.72-5.22)	0.9999	C
Compensated Jaffe-BC AU- Compensated Jaffe-RCc	-7.7	(-11.56-(-3.84))	0.0001	NC
Compensated Jaffe-BC AU- Compensated Jaffe-RCI	-2.99	(-6.57-0.6)	0.1756	C
Compensated Jaffe-BC AU- Compensated Jaffe-RH	-3.49	(-6.77-(-0.22))	0.0222	NC
Compensated Jaffe-BC AU- Compensated Jaffe-SD	0.82	(-3.21-4.85)	0.9994	C
Compensated Jaffe-BC AU- Enzymatic method-BC AU	-2.31	(-6.14-1.51)	0.6074	C
Compensated Jaffe-BC AU- Non-compensated Jaffe-BC AU	-0.75	(-3.64-2.15)	0.9965	C
Compensated Jaffe-BC AU- Non-compensated Jaffe-RH	-3.15	(-7.39-1.09)	0.313	C
Compensated Jaffe-RCc- Compensated Jaffe-RCI	4.72	(-0.02-9.45)	0.0444	NC
Compensated Jaffe- RCc- Compensated Jaffe-RH	4.21	(-0.29-8.71)	0.077	C
Compensated Jaffe-RCc- Compensated Jaffe-SD	8.52	(3.44-13.6)	0.0001	NC
Compensated Jaffe-RCc- Enzymatic method-BC AU	5.39	(0.47-10.31)	0.0163	NC
Compensated Jaffe-RCc- Non-compensated Jaffe-BC AU	6.96	(2.72-11.19)	0.0001	NC
Compensated Jaffe-RCc- Non-compensated Jaffe-RH	4.55	(-0.69-9.79)	0.1347	C
Compensated Jaffe-RCI- Compensated Jaffe-RH	-0.51	(-4.77-3.76)	0.9999	C
Compensated Jaffe-RCI- Compensated Jaffe-SD	3.81	(-1.07-8.68)	0.2495	C
Compensated Jaffe-RCI- Enzymatic method-BC AU	0.67	(-4.03-5.38)	0.9999	C
Compensated Jaffe- RCI- Non-compensated Jaffe-BC AU	2.24	(-1.74-6.22)	0.698	C
Compensated Jaffe- RCI- Non-compensated Jaffe-RH	-0.17	(-5.21-4.88)	0.9999	C
Compensated Jaffe-RH- Compensated Jaffe-SD	4.31	(-0.34-8.96)	0.0829	C
Compensated Jaffe-RH- Enzymatic method - BC AU	1.18	(-3.29-5.65)	0.9959	C
Compensated Jaffe-RH- Non-compensated Jaffe-BC AU	2.75	(-0.96-6.45)	0.3182	C
Compensated Jaffe-RH- Non-compensated Jaffe-RH	0.34	(-4.49-5.17)	0.9999	C
Compensated Jaffe-SD- Enzymatic method-BC AU	-3.13	(-8.19-1.92)	0.5728	C
Compensated Jaffe-SD- Non-compensated Jaffe-BC AU	-1.57	(-5.96-2.82)	0.9702	C
Compensated Jaffe-SD- Non-compensated Jaffe-RH	-3.97	(-9.34-1.4)	0.321	C
Enzymatic method-BC AU- Non-compensated Jaffe-BC AU	1.57	(-2.63-5.77)	0.961	C
Enzymatic method - BC AU- Non-compensated Jaffe-RH	-0.84	(-6.06-4.38)	0.9999	C
Non-compensated Jaffe-BC AU- Non-compensated Jaffe-RH	-2.41	(-6.99-2.17)	0.7704	C

Diff - difference between differences of control and serum sample for each MP pair, Comm – commutability, C – commutable, NC - noncommutable

The overall number of noncommutable MP pairs and percentages of noncommutable decisions per analyte is presented in Table 13.

Table 13. The number of noncommutable decisions for evaluated MP pairs in each EQA survey based on the statistical significance of differences between measurements in control and serum samples

Analyte	1st EQA survey			2nd EQA survey			3rd EQA survey		
	MP pairs	NC	%NC	MP pairs	NC	%NC	MP pairs	NC	%NC
ALT	21	7	33.3	15	5	33.3	21	11	52.4
AP	21	6	28.6	15	4	26.7	21	10	47.6
AMI	10	6	60.0	15	11	73.3	15	/	/
AST	28	/	/	20	12	60.0	20	15	75.0
Calcium	15	/	/	10	/	/	15	2	13.3
Chloride	3	2	66.7	3	3	100	3	2	66.7
Cholesterol	21	9	42.9	15	5	33.3	15	7	46.7
CK	10	2	20.0	10	/	/	15	3	20.0
Creatinine	36	7	19.4	24	7	29.2	36	29	80.6
GGT	15	7	46.7	13	2	15.4	15	10	66.7
Glucose	28	1	3.6	15	4	26.7	21	5	23.8
HDL	21	18	85.7	15	12	80.0	15	11	73.3
Iron	28	2	7.1	21	3	14.3	28	7	25.0
LDH	15	/	/	15	/	/	15	/	/
Phosphate	6	/	/	6	3	50.0	6	/	/
Potassium	10	/	/	10	3	30.0	15	4	26.7
Sodium	15	/	/	15	4	26.7	15	7	46.7
Bilirubin	21	6	28.6	6	3	50.0	15	4	26.7
Proteins	15	/	/	15	2	13.3	15	/	/
Triglycerides	10	1	10.0	10	6	60.0	10	1	10.0
Urate	45	10	22.2	32	3	9.4	36	8	22.2
Urea	21	3	14.3	18	/	/	21	4	19.0
Total	**415**	**87**	**21.0**	**315**	**89**	**28.3**	**388**	**140**	**36.1**

MP-measurement procedure, NC-noncommutable

All three controls show noncommutability for some pairwise combinations of MPs. The number of noncommutable decisions varies depending on the control sample, but even more, depending on the analyte being assessed. The controls are mostly fully or highly commutable for calcium, CK, proteins, and urea. Contrarily, full noncommutability of control samples is described for chloride and HDL. Some of the patterns of noncommutability are related to the control manufacturer, where for example C3/2016 is fully commutable for all MP combinations assessed for measurement of amylase as opposed to noncommutability observed for C1/2016 and C2/2016, which come from the different manufacturer. Quite opposite, for creatinine, controls C1/2016 and C2/2016 are highly commutable, but C3/2016 is noncommutable for almost 81% of all MP combinations assessed. The commutability of controls for some analytes might depend on concentrations assessed, with opposite conclusions on overall commutability for normal and pathological concentrations levels. This is observed for triglycerides and GGT, and to a lower extent for some other analytes like urea and urate. For most analytes, the commutability does not seem to be connected to measured concentrations. Overall, controls C1/2016 and C2/2016 show somewhat better commutability with MP combinations used in this EQA than C3/2016. When mostly normal-concentration level controls are compared (C1/2016 and C3/2016) from different manufacturers, the control C1/2016 shows higher overall commutability, with 87 noncommutable pairwise combinations of MPs as opposed to 140 noncommutable results in C3/2016.

4.2.2 False flagging method in evaluating commutability

According to written instructions sent to each participant, the laboratories measured both control and serum samples in the same run on the instrument. The results obtained from each laboratory were grouped together and accordingly MP groups were formed as listed in Table 3. Based on the analysis of both serum and control samples for each MP, the rate of falsely flagged laboratories according to the consensus mean and predefined APS of CROQALM is assessed, according to the procedure described in Materials and methods section 3.2.

Harmonisation between MPs is calculated as the probability of changing the flagging status of laboratories above the threshold limit of 20% when the results on serum samples of two MPs groups are joined into one group. The value of 100% indicates perfect harmonisation, whereas

the value of 0% indicates a total lack of harmonisation. The value above 95% is considered as satisfactory proof of harmonisation.

Commutability is calculated as the difference between flagging rates of laboratories using results of measurement on control serum compared to the measurement results obtained on the serum sample. The limit of 20% is set as the maximum allowable false flagging rate of laboratories due to noncommutability (3.5 Materials and Methods). Percentage commutability is calculated as a number of bootstrapped results not exceeding the predefined limit of 20%. Control samples are defined as commutable for evaluated MP pair if percentage commutability is ≥ 95%.

4.2.2.1 Commutability evaluation of control sample C1/2016 using the false flagging method

The results of the analysis of lyophilised control sample C1/2016 by applying the false flagging method on results obtained from participating laboratories in CROQALM EQA survey 1 are presented in Table 14. The mean values of each MP for serum samples and lyophilised control samples are presented in the same table, together with the results for harmonisation between evaluated MP pairs. The results are accompanied by a contingency table showing the number of harmonised and nonharmonised MPs resulting in (non)commutability (Table 15).

Table 14. The results of commutability evaluation of EQA control sample C1/2016 using the false flagging method

		EQA SURVEY 1					
MP 1	**MP 2**	**Mean MP1 (serum)**	**Mean MP2 (serum)**	**Mean MP1 (control)**	**Mean MP2 (control)**	**% harmo-nisation**	**% commu-tability**
			ALT				
IFCC- BC AU	IFCC- SD	34.42	39.09	45.3	47.91	15.3	54.9
IFCC- BC AU	Photometry UV- AA	34.42	32.25	45.3	43.62	100	100
IFCC- BC AU	Photometry UV- BC AU	34.42	32.94	45.3	45.3	100	100
IFCC- BC AU	Photometry UV- RCc	34.42	30.86	45.3	43.5	96.8	96.8

EQA SURVEY 1

IFCC- BC AU	Photometry UV- RCI	34.42	30.07	45.3	42	99.8	99.8
IFCC- BC AU	Photometry UV- RH	34.42	31.88	45.3	43.72	100	100
IFCC- SD	Photometry UV- AA	39.09	32.25	47.91	43.62	11.8	71.1
IFCC- SD	Photometry UV- BC AU	39.09	32.94	47.91	45.3	1.2	20.3
IFCC- SD	Photometry UV- RCc	39.09	30.86	47.91	43.5	3	29.8
IFCC- SD	Photometry UV- RCI	39.09	30.07	47.91	42	0.2	37.3
IFCC- SD	Photometry UV- RH	39.09	31.88	47.91	43.72	1.3	40.2
Photometry UV- AA	Photometry UV- BC AU	32.25	32.94	43.62	45.3	100	100
Photometry UV- AA	Photometry UV- RCc	32.25	30.86	43.62	43.5	100	100
Photometry UV- AA	Photometry UV- RCI	32.25	30.07	43.62	42	100	100
Photometry UV- AA	Photometry UV- RH	32.25	31.88	43.62	43.72	100	100
Photometry UV- BC AU	Photometry UV- RCc	32.94	30.86	45.3	43.5	100	100
Photometry UV- BC AU	Photometry UV- RCI	32.94	30.07	45.3	42	100	100
Photometry UV- BC AU	Photometry UV- RH	32.94	31.88	45.3	43.72	100	100
Photometry UV- RCc	Photometry UV- RCI	30.86	30.07	43.5	42	100	100
Photometry UV- RCc	Photometry UV- RH	30.86	31.88	43.5	43.72	100	100
Photometry UV- RCI	Photometry UV- RH	30.07	31.88	42	43.72	100	100
			ALP				
IFCC- AA	IFCC- BC AU	73	75.14	107.1	107.62	100	99.2
IFCC- AA	IFCC- HP	73	79.33	107.1	108.67	91.2	95.1
IFCC- AA	IFCC- RCc	73	71.69	107.1	97.77	100	82.1
IFCC- AA	IFCC- RCI	73	74.5	107.1	99.82	100	94.9
IFCC- AA	IFCC- RH	73	72.88	107.1	97.88	100	43.7
IFCC- AA	IFCC - SD	73	74.4	107.1	102.5	100	100
IFCC- BC AU	IFCC- HP	75.14	79.33	107.62	108.67	87.2	93.3
IFCC- BC AU	IFCC- RCc	75.14	71.69	107.62	97.77	100	54.3
IFCC- BC AU	IFCC- RCI	75.14	74.5	107.62	99.82	100	99.3
IFCC- BC AU	IFCC- RH	75.14	72.88	107.62	97.88	100	18.8
IFCC- BC AU	IFCC - SD	75.14	74.4	107.62	102.5	100	100
IFCC- HP	IFCC- RCc	79.33	71.69	108.67	97.77	77.6	67.8
IFCC- HP	IFCC- RCI	79.33	74.5	108.67	99.82	93.5	73.4
IFCC- HP	IFCC- RH	79.33	72.88	108.67	97.88	71.5	51.6
IFCC- HP	IFCC - SD	79.33	74.4	108.67	102.5	97.7	99
IFCC- RCc	IFCC- RCI	71.69	74.5	97.77	99.82	100	100
IFCC- RCc	IFCC- RH	71.69	72.88	97.77	97.88	100	100
IFCC- RCc	IFCC - SD	71.69	74.4	97.77	102.5	100	100

IFCC- RCI	IFCC- RH	74.5	72.88	99.82	97.88	100	100
IFCC- RCI	IFCC - SD	74.5	74.4	99.82	102.5	100	100
IFCC- RH	IFCC - SD	72.88	74.4	97.88	102.5	100	100
			AMY				
IFCC- AA	IFCC- BC AU	61.78	65.89	98.11	101.53	91.1	91.8
IFCC- AA	IFCC- RCc	61.78	64.85	98.11	98.23	99.8	99.9
IFCC- AA	IFCC- RCI	61.78	66.36	98.11	99.87	98.8	98.8
IFCC- AA	IFCC- RH	61.78	63	98.11	96.73	99.9	99.9
IFCC- BC AU	IFCC- RCc	65.89	64.85	101.53	98.23	100	100
IFCC- BC AU	IFCC- RCI	65.89	66.36	101.53	99.87	100	100
IFCC- BC AU	IFCC- RH	65.89	63	101.53	96.73	100	100
IFCC- RCc	IFCC- RCI	64.85	66.36	98.23	99.87	100	100
IFCC- RCc	IFCC- RH	64.85	63	98.23	96.73	100	100
IFCC- RCI	IFCC- RH	66.36	63	99.87	96.73	100	100
			AST				
IFCC- BC AU	IFCC- RH	30.31	27.33	43.52	42.5	94.2	95.3
IFCC- BC AU	IFCC- SD	30.31	29.2	43.52	43.18	100	100
IFCC- BC AU	Photometry UV- AA	30.31	26	43.52	40	99.8	99.8
IFCC- BC AU	Photometry UV- BC AU	30.31	29.94	43.52	43.25	100	100
IFCC- BC AU	Photometry UV- RCc	30.31	25.71	43.52	39.75	99.9	99.9
IFCC- BC AU	Photometry UV- RCI	30.31	26.71	43.52	39.5	100	100
IFCC- BC AU	Photometry UV- RH	30.31	26.94	43.52	40.17	100	100
IFCC- RH	IFCC- SD	27.33	29.2	42.5	43.18	100	100
IFCC- RH	Photometry UV- AA	27.33	26	42.5	40	100	100
IFCC- RH	Photometry UV- BC AU	27.33	29.94	42.5	43.25	96.1	97.1
IFCC- RH	Photometry UV- RCc	27.33	25.71	42.5	39.75	100	100
IFCC- RH	Photometry UV- RCI	27.33	26.71	42.5	39.5	100	100
IFCC- RH	Photometry UV- RH	27.33	26.94	42.5	40.17	100	99.8
IFCC- SD	Photometry UV- AA	29.2	26	43.18	40	100	100
IFCC- SD	Photometry UV- BC AU	29.2	29.94	43.18	43.25	100	100
IFCC- SD	Photometry UV- RCc	29.2	25.71	43.18	39.75	100	100
IFCC- SD	Photometry UV- RCI	29.2	26.71	43.18	39.5	100	100
IFCC- SD	Photometry UV- RH	29.2	26.94	43.18	40.17	100	100
Photometry UV- AA	Photometry UV- BC AU	26	29.94	40	43.25	100	100
Photometry UV- AA	Photometry UV- RCc	26	25.71	40	39.75	100	100
Photometry UV- AA	Photometry UV- RCI	26	26.71	40	39.5	100	100

EQA SURVEY 1

Photometry UV- AA	Photometry UV- RH	26	26.94	40	40.17	100	100
Photometry UV- BC AU	Photometry UV- RCc	29.94	25.71	43.25	39.75	100	100
Photometry UV- BC AU	Photometry UV- RCI	29.94	26.71	43.25	39.5	100	100
Photometry UV- BC AU	Photometry UV- RH	29.94	26.94	43.25	40.17	100	100
Photometry UV- RCc	Photometry UV- RCI	25.71	26.71	39.75	39.5	100	100
Photometry UV- RCc	Photometry UV- RH	25.71	26.94	39.75	40.17	100	100
Photometry UV- RCI	Photometry UV- RH	26.71	26.94	39.5	40.17	100	100
CALCIUM							
Asenaso III- AA	Asenaso III- BC AU	2.31	2.31	2.35	2.35	100	99.9
Asenaso III- AA	NM-BAPTA- RCI	2.31	2.26	2.35	2.35	100	100
Asenaso III- AA	cresolphthalein - BC AU	2.31	2.29	2.35	2.32	100	95.6
Asenaso III- AA	cresolphthalein - RCI	2.31	2.31	2.35	2.38	100	100
Asenaso III- AA	cresolphthalein - SD	2.31	2.23	2.35	2.29	98.4	99.4
Asenaso III- BC AU	NM-BAPTA- RCI	2.31	2.26	2.35	2.35	89.5	92.2
Asenaso III- BC AU	cresolphthalein - BC AU	2.31	2.29	2.35	2.32	100	83.8
Asenaso III- BC AU	cresolphthalein - RCI	2.31	2.31	2.35	2.38	100	97.8
Asenaso III- BC AU	cresolphthalein - SD	2.31	2.23	2.35	2.29	28.7	56
NM-BAPTA- RCI	cresolphthalein - BC AU	2.26	2.29	2.35	2.32	100	99.4
NM-BAPTA- RCI	cresolphthalein - RCI	2.26	2.31	2.35	2.38	99.9	100
NM-BAPTA- RCI	cresolphthalein - SD	2.26	2.23	2.35	2.29	100	99.8
cresolphthalein - BC AU	cresolphthalein - RCI	2.29	2.31	2.32	2.38	100	86
cresolphthalein - BC AU	cresolphthalein - SD	2.29	2.23	2.32	2.29	97.7	99.1
cresolphthalein - RCI	cresolphthalein - SD	2.31	2.23	2.38	2.29	92	96.6
CHLORIDE							
Indirect ISE - AA	Indirect ISE - BC AU	103.5	103.35	114.8	114.61	97.9	99
Indirect ISE - AA	Indirect ISE - SD	103.5	101	114.8	110	100	91.2
Indirect ISE - BC AU	Indirect ISE - SD	103.35	101	114.61	110	100	12.8
CHOLESTEROL							
CHOD-PAP - AA	CHOD-PAP - BC AU	6.13	6.39	4.14	4.28	100	100

CHOD-PAP - AA	CHOD-PAP - HP	6.13	5.98	4.14	4.03	100	100
CHOD-PAP - AA	CHOD-PAP - RCc	6.13	6.02	4.14	4.06	100	100
CHOD-PAP - AA	CHOD-PAP - RCI	6.13	6.17	4.14	4.21	100	100
CHOD-PAP - AA	CHOD-PAP - RH	6.13	6.06	4.14	4.11	100	100
CHOD-PAP - AA	CHOD-PAP - SD	6.13	6.17	4.14	3.72	100	15.8
CHOD-PAP - BC AU	CHOD-PAP - HP	6.39	5.98	4.28	4.03	97.4	98.6
CHOD-PAP - BC AU	CHOD-PAP - RCc	6.39	6.02	4.28	4.06	99.2	99.5
CHOD-PAP - BC AU	CHOD-PAP - RCI	6.39	6.17	4.28	4.21	100	100
CHOD-PAP - BC AU	CHOD-PAP - RH	6.39	6.06	4.28	4.11	100	100
CHOD-PAP - BC AU	CHOD-PAP - SD	6.39	6.17	4.28	3.72	100	0.2
CHOD-PAP - HP	CHOD-PAP - RCc	5.98	6.02	4.03	4.06	100	100
CHOD-PAP - HP	CHOD-PAP - RCI	5.98	6.17	4.03	4.21	100	100
CHOD-PAP - HP	CHOD-PAP - RH	5.98	6.06	4.03	4.11	100	100
CHOD-PAP - HP	CHOD-PAP - SD	5.98	6.17	4.03	3.72	100	80.4
CHOD-PAP - RCc	CHOD-PAP - RCI	6.02	6.17	4.06	4.21	100	100
CHOD-PAP - RCc	CHOD-PAP - RH	6.02	6.06	4.06	4.11	100	100
CHOD-PAP - RCc	CHOD-PAP - SD	6.02	6.17	4.06	3.72	100	26.8
CHOD-PAP - RCI	CHOD-PAP - RH	6.17	6.06	4.21	4.11	100	100
CHOD-PAP - RCI	CHOD-PAP - SD	6.17	6.17	4.21	3.72	100	2
CHOD-PAP - RH	CHOD-PAP - SD	6.06	6.17	4.11	3.72	100	7.3
			CK				
IFCC- AA	IFCC- BC AU	203.7	210.59	100.1	102.16	100	100
IFCC- AA	IFCC- RCc	203.7	213.33	100.1	102.17	100	100
IFCC- AA	IFCC- RCI	203.7	208	100.1	102.18	100	100
IFCC- AA	IFCC- RH	203.7	203.56	100.1	102.56	100	100
IFCC- BC AU	IFCC- RCc	210.59	213.33	102.16	102.17	100	100
IFCC- BC AU	IFCC- RCI	210.59	208	102.16	102.18	100	100
IFCC- BC AU	IFCC- RH	210.59	203.56	102.16	102.56	100	100
IFCC- RCc	IFCC- RCI	213.33	208	102.17	102.18	100	100
IFCC- RCc	IFCC- RH	213.33	203.56	102.17	102.56	100	100
IFCC- RCI	IFCC- RH	208	203.56	102.18	102.56	100	100
			CREATININE				
Compensated Jaffe - AA	Compensated Jaffe - BC AU	75.5	75.6	79.5	76.98	99.4	97.2
Compensated Jaffe - AA	Compensated Jaffe - RCc	75.5	75.2	79.5	84.5	100	98

EQA SURVEY 1

Compensated Jaffe - AA	Compensated Jaffe - RCI	75.5	78	79.5	82.58	99.4	99.5
Compensated Jaffe - AA	Compensated Jaffe - RH	75.5	80.07	79.5	85.5	92.9	91.9
Compensated Jaffe - AA	Compensated Jaffe - SD	75.5	77.56	79.5	78.33	100	100
Compensated Jaffe - AA	Enzymatic method - BC AU	75.5	74.82	79.5	79.1	100	100
Compensated Jaffe - AA	Non-compensated Jaffe - BC AU	75.5	86.05	79.5	88.45	0.5	55.4
Compensated Jaffe - AA	Non-compensated Jaffe - RH	75.5	79.25	79.5	84	99.1	99.2
Compensated Jaffe - BC AU	Compensated Jaffe - RCc	75.6	75.2	76.98	84.5	100	22.8
Compensated Jaffe - BC AU	Compensated Jaffe - RCI	75.6	78	76.98	82.58	97.8	72.6
Compensated Jaffe - BC AU	Compensated Jaffe - RH	75.6	80.07	76.98	85.5	99.8	28.4
Compensated Jaffe - BC AU	Compensated Jaffe - SD	75.6	77.56	76.98	78.33	99	99.1
Compensated Jaffe - BC AU	Enzymatic method - BC AU	75.6	74.82	76.98	79.1	100	100
Compensated Jaffe - BC AU	Non-compensated Jaffe - BC AU	75.6	86.05	76.98	88.45	4.7	78.3
Compensated Jaffe - BC AU	Non-compensated Jaffe - RH	75.6	79.25	76.98	84	82.5	45.6
Compensated Jaffe - RCc	Compensated Jaffe - RCI	75.2	78	84.5	82.58	100	100
Compensated Jaffe - RCc	Compensated Jaffe - RH	75.2	80.07	84.5	85.5	97.5	99.4
Compensated Jaffe - RCc	Compensated Jaffe - SD	75.2	77.56	84.5	78.33	100	99.3
Compensated Jaffe - RCc	Enzymatic method - BC AU	75.2	74.82	84.5	79.1	100	99.4
Compensated Jaffe - RCc	Non-compensated Jaffe - BC AU	75.2	86.05	84.5	88.45	0.1	5.2
Compensated Jaffe - RCc	Non-compensated Jaffe - RH	75.2	79.25	84.5	84	97.3	99.4
Compensated Jaffe - RCI	Compensated Jaffe - RH	78	80.07	82.58	85.5	99.2	99.4
Compensated Jaffe - RCI	Compensated Jaffe - SD	78	77.56	82.58	78.33	100	100
Compensated Jaffe - RCI	Enzymatic method - BC AU	78	74.82	82.58	79.1	99.8	100
Compensated Jaffe - RCI	Non-compensated Jaffe - BC AU	78	86.05	82.58	88.45	22.3	81.3

Compensated Jaffe - RCI	Non-compensated Jaffe - RH	78	79.25	82.58	84	99.8	98.9
Compensated Jaffe - RH	Compensated Jaffe - SD	80.07	77.56	85.5	78.33	100	99
Compensated Jaffe - RH	Enzymatic method - BC AU	80.07	74.82	85.5	79.1	98.8	100
Compensated Jaffe - RH	Non-compensated Jaffe - BC AU	80.07	86.05	85.5	88.45	100	100
Compensated Jaffe - RH	Non-compensated Jaffe - RH	80.07	79.25	85.5	84	100	99.8
Compensated Jaffe - SD	Enzymatic method - BC AU	77.56	74.82	78.33	79.1	99.9	99.9
Compensated Jaffe - SD	Non-compensated Jaffe - BC AU	77.56	86.05	78.33	88.45	11.3	82.4
Compensated Jaffe - SD	Non-compensated Jaffe - RH	77.56	79.25	78.33	84	100	91.3
Enzymatic method - BC AU	Non-compensated Jaffe - BC AU	74.82	86.05	79.1	88.45	0.5	28.4
Enzymatic method - BC AU	Non-compensated Jaffe - RH	74.82	79.25	79.1	84	95.2	97.6
Non-compensated Jaffe - BC AU	Non-compensated Jaffe - RH	86.05	79.25	88.45	84	38.4	90.3
			GGT				
IFCC- AA	IFCC- BC AU	47.59	46.59	36.09	36.38	100	100
IFCC- AA	IFCC- HP	47.59	44.62	36.09	36.01	99.9	100
IFCC- AA	IFCC- RCc	47.59	47.21	36.09	35.99	100	100
IFCC- AA	IFCC- RCI	47.59	47.29	36.09	36.13	100	100
IFCC- AA	IFCC- RH	47.59	47.23	36.09	36.26	100	100
IFCC- AA	IFCC- SD	47.59	48.32	36.09	39.28	100	92.7
IFCC- BC AU	IFCC- HP	46.59	44.62	36.38	36.01	98.4	98.2
IFCC- BC AU	IFCC- RCc	46.59	47.21	36.38	35.99	100	100
IFCC- BC AU	IFCC- RCI	46.59	47.29	36.38	36.13	100	100
IFCC- BC AU	IFCC- RH	46.59	47.23	36.38	36.26	100	100
IFCC- BC AU	IFCC- SD	46.59	48.32	36.38	39.28	100	37.4
IFCC- HP	IFCC- RCc	44.62	47.21	36.01	35.99	99.9	100
IFCC- HP	IFCC- RCI	44.62	47.29	36.01	36.13	99.8	99.9
IFCC- HP	IFCC- RH	44.62	47.23	36.01	36.26	99.2	99.3
IFCC- HP	IFCC- SD	44.62	48.32	36.01	39.28	99.3	75.5
IFCC- RCc	IFCC- RCI	47.21	47.29	35.99	36.13	100	100
IFCC- RCc	IFCC- RH	47.21	47.23	35.99	36.26	100	100
IFCC- RCc	IFCC- SD	47.21	48.32	35.99	39.28	100	89.3
IFCC- RCI	IFCC- RH	47.29	47.23	36.13	36.26	100	100
IFCC- RCI	IFCC- SD	47.29	48.32	36.13	39.28	100	77.2
IFCC- RH	IFCC- SD	47.23	48.32	36.26	39.28	100	60.8

EQA SURVEY 1

GLUCOSE							
GOD-PAP - BC AU	GOD-PAP - HP	5.31	5.25	4.43	4.43	100	100
GOD-PAP - BC AU	GOD-PAP - RH	5.31	5.13	4.43	4.36	100	100
GOD-PAP - BC AU	Hexokinase - AA	5.31	5.22	4.43	4.41	100	100
GOD-PAP - BC AU	Hexokinase - BC AU	5.31	5.27	4.43	4.39	100	100
GOD-PAP - BC AU	Hexokinase - RCc	5.31	5.19	4.43	4.34	100	100
GOD-PAP - BC AU	Hexokinase - RCI	5.31	5.21	4.43	4.37	100	100
GOD-PAP - BC AU	Hexokinase - SD	5.31	5.33	4.43	4.52	100	100
GOD-PAP - HP	GOD-PAP - RH	5.25	5.13	4.43	4.36	99.1	99.4
GOD-PAP - HP	Hexokinase - AA	5.25	5.22	4.43	4.41	100	100
GOD-PAP - HP	Hexokinase - BC AU	5.25	5.27	4.43	4.39	100	100
GOD-PAP - HP	Hexokinase - RCc	5.25	5.19	4.43	4.34	100	100
GOD-PAP - HP	Hexokinase - RCI	5.25	5.21	4.43	4.37	100	100
GOD-PAP - HP	Hexokinase - SD	5.25	5.33	4.43	4.52	100	99.7
GOD-PAP - RH	Hexokinase - AA	5.13	5.22	4.36	4.41	100	100
GOD-PAP - RH	Hexokinase - BC AU	5.13	5.27	4.36	4.39	99.7	99.9
GOD-PAP - RH	Hexokinase - RCc	5.13	5.19	4.36	4.34	100	100
GOD-PAP - RH	Hexokinase - RCI	5.13	5.21	4.36	4.37	100	100
GOD-PAP - RH	Hexokinase - SD	5.13	5.33	4.36	4.52	92.5	99.2
Hexokinase - AA	Hexokinase - BC AU	5.22	5.27	4.41	4.39	100	100
Hexokinase - AA	Hexokinase - RCc	5.22	5.28	4.41	4.34	100	100
Hexokinase - AA	Hexokinase - RCI	5.22	5.29	4.41	4.37	100	100
Hexokinase - AA	Hexokinase - SD	5.22	5.30	4.41	4.52	100	100
Hexokinase - BC AU	Hexokinase - RCc	5.27	5.31	4.39	4.34	100	100
Hexokinase - BC AU	Hexokinase - RCI	5.27	5.32	4.39	4.37	100	100
Hexokinase - BC AU	Hexokinase - SD	5.27	5.33	4.39	4.52	100	93.3
Hexokinase - RCc	Hexokinase - RCI	5.19	5.34	4.34	4.37	100	100
Hexokinase - RCc	Hexokinase - SD	5.19	5.35	4.34	4.52	100	99.9
Hexokinase - RCI	Hexokinase - SD	5.21	5.36	4.37	4.52	100	100

EQA SURVEY 1

HDL							
Homogenous - AA	Homogenous - BC AU	1.71	1.65	1.26	1.1	100	1.1
Homogenous - AA	Homogenous - HP	1.71	1.65	1.26	1.33	100	75.2
Homogenous - AA	Homogenous - RCc	1.71	1.78	1.26	1.05	100	8.9
Homogenous - AA	Homogenous - RCI	1.71	1.71	1.26	1.02	100	0
Homogenous - AA	Homogenous - RH	1.71	1.75	1.26	1.18	100	76.9
Homogenous - AA	Homogenous - SD	1.71	1.78	1.26	1.07	99.4	27.1
Homogenous - BC AU	Homogenous - HP	1.65	1.65	1.1	1.33	100	0.1
Homogenous - BC AU	Homogenous - RCc	1.65	1.78	1.1	1.05	39.5	81.8
Homogenous - BC AU	Homogenous - RCI	1.65	1.71	1.1	1.02	100	93.6
Homogenous - BC AU	Homogenous - RH	1.65	1.75	1.1	1.18	93.2	99.2
Homogenous - BC AU	Homogenous - SD	1.65	1.78	1.1	1.07	39.4	57.5
Homogenous - HP	Homogenous - RCc	1.65	1.78	1.33	1.05	99.8	0.4
Homogenous - HP	Homogenous - RCI	1.65	1.71	1.33	1.02	100	0
Homogenous - HP	Homogenous - RH	1.65	1.75	1.33	1.18	100	16.3
Homogenous - HP	Homogenous - SD	1.65	1.78	1.33	1.07	100	1.5
Homogenous - RCc	Homogenous - RCI	1.78	1.71	1.05	1.02	98.6	99.2
Homogenous - RCc	Homogenous - RH	1.78	1.75	1.05	1.18	99.6	47
Homogenous - RCc	Homogenous - SD	1.78	1.78	1.05	1.07	99.8	99.3
Homogenous - RCI	Homogenous - RH	1.71	1.75	1.02	1.18	100	75
Homogenous - RCI	Homogenous - SD	1.71	1.78	1.02	1.07	98.5	98.9
Homogenous - RH	Homogenous - SD	1.75	1.78	1.18	1.07	99.5	82.4
IRON							
Ferene - AA	Ferene - HP	20.43	20.5	17.71	16.86	100	100
Ferene - AA	Ferene - RH	20.43	21.08	17.71	17.92	100	100
Ferene - AA	Ferene - SD	20.43	20.2	17.71	17.2	100	100
Ferene - AA	Ferrozine - RCc	20.43	21.25	17.71	17.92	100	100
Ferene - AA	Ferrozine - RCI	20.43	21.47	17.71	18.6	100	100
Ferene - AA	Ferrozine - RH	20.43	21.36	17.71	18.09	100	100
Ferene - AA	TPTZ - BC AU	20.43	20.96	17.71	17.43	100	100
Ferene - HP	Ferene - RH	20.5	21.08	16.86	17.92	100	100
Ferene - HP	Ferene - SD	20.5	20.2	16.86	17.2	100	100

Ferene - HP	Ferrozine - RCc	20.5	21.25	16.86	17.92	100	100
Ferene - HP	Ferrozine - RCI	20.5	21.47	16.86	18.6	100	100
Ferene - HP	Ferrozine - RH	20.5	21.36	16.86	18.09	100	100
Ferene - HP	TPTZ - BC AU	20.5	20.96	16.86	17.43	100	100
Ferene - RH	Ferene - SD	21.08	20.2	17.92	17.2	100	100
Ferene - RH	Ferrozine - RCc	21.08	21.25	17.92	17.92	100	100
Ferene - RH	Ferrozine - RCI	21.08	21.47	17.92	18.6	100	100
Ferene - RH	Ferrozine - RH	21.08	21.36	17.92	18.09	100	100
Ferene - RH	TPTZ - BC AU	21.08	20.96	17.92	17.43	100	100
Ferene - SD	Ferrozine - RCc	20.2	21.25	17.2	17.92	100	100
Ferene - SD	Ferrozine - RCI	20.2	21.47	17.2	18.6	100	100
Ferene - SD	Ferrozine - RH	20.2	21.36	17.2	18.09	100	100
Ferene - SD	TPTZ - BC AU	20.2	20.96	17.2	17.43	100	100
Ferrozine - RCc	Ferrozine - RCI	21.25	21.47	17.92	18.6	100	100
Ferrozine - RCc	Ferrozine - RH	21.25	21.36	17.92	18.09	100	100
Ferrozine - RCc	TPTZ - BC AU	21.25	20.96	17.92	17.43	100	100
Ferrozine - RCI	Ferrozine - RH	21.47	21.36	18.6	18.09	100	100
Ferrozine - RCI	TPTZ - BC AU	21.47	20.96	18.6	17.43	100	100
Ferrozine - RH	TPTZ - BC AU	21.36	20.96	18.09	17.43	100	100
LDH							
IFCC- AA	IFCC- BC AU	177.89	176.85	144.11	145.42	100	100
IFCC- AA	IFCC- RCc	177.89	170.71	144.11	139.57	100	100
IFCC- AA	IFCC- RCI	177.89	183.62	144.11	152.31	100	100
IFCC- AA	IFCC- RH	177.89	175.5	144.11	148	100	100
IFCC- AA	IFCC - SD	177.89	176.83	144.11	148.17	100	100
IFCC- BC AU	IFCC- RCc	176.85	170.71	145.42	139.57	100	100
IFCC- BC AU	IFCC- RCI	176.85	183.62	145.42	152.31	100	100
IFCC- BC AU	IFCC- RH	176.85	175.5	145.42	148	100	100
IFCC- BC AU	IFCC - SD	176.85	176.83	145.42	148.17	100	97.8
IFCC- RCc	IFCC- RCI	170.71	183.62	139.57	152.31	100	100
IFCC- RCc	IFCC- RH	170.71	175.5	139.57	148	100	100
IFCC- RCc	IFCC - SD	170.71	176.83	139.57	148.17	100	98.9
IFCC- RCI	IFCC- RH	183.62	175.5	152.31	148	100	100
IFCC- RCI	IFCC - SD	183.62	176.83	152.31	148.17	100	100
IFCC- RH	IFCC - SD	175.5	176.83	148	148.17	100	100
PHOSPHATE							
Ammonium-molybdate - AA	Ammonium-molybdate - BC AU	0.99	0.98	1.05	1.03	100	99.7

Ammonium-molybdate - AA	Ammonium-molybdate - RCc	0.99	0.96	1.05	1.03	100	100
Ammonium-molybdate - AA	Ammonium-molybdate - RCI	0.99	1	1.05	1.07	100	100
Ammonium-molybdate - BC AU	Ammonium-molybdate - RCc	0.98	0.96	1.03	1.03	100	100
Ammonium-molybdate - BC AU	Ammonium-molybdate - RCI	0.98	1	1.03	1.07	100	100
Ammonium-molybdate - RCc	Ammonium-molybdate - RCI	0.96	1	1.03	1.07	100	100
POTASSIUM							
FES- CC	Indirect ISE-AA	4.11	4.07	3.65	3.53	100	88
FES- CC	Indirect ISE-BC AU	4.11	4.07	3.65	3.57	100	87.5
FES- CC	Indirect ISE-RCc	4.11	4.19	3.65	3.65	100	100
FES- CC	Indirect ISE-RCI	4.11	4.13	3.65	3.59	100	99.9
FES- CC	Indirect ISE-SD	4.11	4.03	3.65	3.51	100	99.9
Indirect ISE-AA	Indirect ISE-BC AU	4.07	4.07	3.53	3.57	100	98.9
Indirect ISE-AA	Indirect ISE-RCc	4.07	4.19	3.53	3.65	100	99.2
Indirect ISE-AA	Indirect ISE-RCI	4.07	4.13	3.53	3.59	100	100
Indirect ISE-AA	Indirect ISE-SD	4.07	4.03	3.53	3.51	100	100
Indirect ISE-BC AU	Indirect ISE-RCc	4.07	4.19	3.57	3.65	99.7	99.7
Indirect ISE-BC AU	Indirect ISE-RCI	4.07	4.13	3.57	3.59	100	100
Indirect ISE-BC AU	Indirect ISE-SD	4.07	4.03	3.57	3.51	100	100
Indirect ISE-RCc	Indirect ISE-RCI	4.19	4.13	3.65	3.59	100	100
Indirect ISE-RCc	Indirect ISE-SD	4.19	4.03	3.65	3.51	100	100
Indirect ISE-RCI	Indirect ISE-SD	4.13	4.03	3.59	3.51	100	100
SODIUM							
FES - CC	Indirect ISE - AA	140.36	139.78	138.95	138.44	100	100
FES - CC	Indirect ISE - BC AU	140.36	138.88	138.95	137.9	100	100
FES - CC	Indirect ISE - RCc	140.36	139.11	138.95	138.1	100	100
FES - CC	Indirect ISE - RCI	140.36	140	138.95	139.7	100	100
FES - CC	Indirect ISE - SD	140.36	139.3	138.95	139.3	100	100

Indirect ISE - AA	Indirect ISE - BC AU	139.78	138.88	138.44	137.9	100	100
Indirect ISE - AA	Indirect ISE - RCc	139.78	139.11	138.44	138.1	100	100
Indirect ISE - AA	Indirect ISE - RCI	139.78	140	138.44	139.7	100	100
Indirect ISE - AA	Indirect ISE - SD	139.78	139.3	138.44	139.3	100	100
Indirect ISE - BC AU	Indirect ISE - RCc	138.88	139.11	137.9	138.1	100	100
Indirect ISE - BC AU	Indirect ISE - RCI	138.88	140	137.9	139.7	100	100
Indirect ISE - BC AU	Indirect ISE - SD	138.88	139.3	137.9	139.3	100	100
Indirect ISE - RCc	Indirect ISE - RCI	139.11	140	138.1	139.7	100	100
Indirect ISE - RCc	Indirect ISE - SD	139.11	139.3	138.1	139.3	100	100
Indirect ISE - RCI	Indirect ISE - SD	140	139.3	139.7	139.3	100	100
BILIRUBIN							
Diazo - AA	Diazo - BC AU	7.12	7.58	18.38	19.18	100	100
Diazo - AA	Diazo - HP	7.12	6.86	18.38	18.29	100	100
Diazo - AA	Diazo - RCc	7.12	5.75	18.38	16.67	100	100
Diazo - AA	Diazo - RCI	7.12	6.06	18.38	16.44	100	100
Diazo - AA	Diazo - RH	7.12	6.95	18.38	17.82	100	100
Diazo - AA	Diazo - SD	7.12	5.92	18.38	16.83	100	100
Diazo - BC AU	Diazo - HP	7.58	6.86	19.18	18.29	100	100
Diazo - BC AU	Diazo - RCc	7.58	5.75	19.18	16.67	99	77.6
Diazo - BC AU	Diazo - RCI	7.58	6.06	19.18	16.44	100	85.1
Diazo - BC AU	Diazo - RH	7.58	6.95	19.18	17.82	100	100
Diazo - BC AU	Diazo - SD	7.58	5.92	19.18	16.83	100	69
Diazo - HP	Diazo - RCc	6.86	5.75	18.29	16.67	100	100
Diazo - HP	Diazo - RCI	6.86	6.06	18.29	16.44	100	100
Diazo - HP	Diazo - RH	6.86	6.95	18.29	17.82	100	100
Diazo - HP	Diazo - SD	6.86	5.92	18.29	16.83	100	100
Diazo - RCc	Diazo - RCI	5.75	6.06	16.67	16.44	100	100
Diazo - RCc	Diazo - RH	5.75	6.95	16.67	17.82	100	100
Diazo - RCc	Diazo - SD	5.75	5.92	16.67	16.83	100	100
Diazo - RCI	Diazo - RH	6.06	6.95	16.44	17.82	100	100
Diazo - RCI	Diazo - SD	6.06	5.92	16.44	16.83	100	100
Diazo - RH	Diazo - SD	6.95	5.92	17.82	16.83	100	100
PROTEINS							
Biuret - AA	Biuret - BC AU	71.12	72.58	66.22	68.06	99.4	83.6
Biuret - AA	Biuret - RCc	71.12	72.69	66.22	67.67	100	99.5
Biuret - AA	Biuret - RCI	71.12	71.73	66.22	67.2	100	99.2
Biuret - AA	Biuret - RH	71.12	73.06	66.22	68.06	99.8	96.3
Biuret - AA	Biuret - SD	71.12	73.38	66.22	68.5	100	99.2
Biuret - BC AU	Biuret - RCc	72.58	72.69	68.06	67.67	100	100
Biuret - BC AU	Biuret - RCI	72.58	71.73	68.06	67.2	100	100
Biuret - BC AU	Biuret - RH	72.58	73.06	68.06	68.06	100	100

Biuret - BC AU	Biuret - SD	72.58	73.38	68.06	68.5	100	100
Biuret - RCc	Biuret - RCI	72.69	71.73	67.67	67.2	100	100
Biuret - RCc	Biuret - RH	72.69	73.06	67.67	68.06	100	100
Biuret - RCc	Biuret - SD	72.69	73.38	67.67	68.5	100	100
Biuret - RCI	Biuret - RH	71.73	73.06	67.2	68.06	100	100
Biuret - RCI	Biuret - SD	71.73	73.38	67.2	68.5	100	100
Biuret - RH	Biuret - SD	73.06	73.38	68.06	68.5	100	100
TRIGLYCERIDES							
GPO-PAP - AA	GPO-PAP - BC AU	1.08	1.11	1.1	1.1	100	100
GPO-PAP - AA	GPO-PAP - RCc	1.08	1.04	1.1	1.08	100	100
GPO-PAP - AA	GPO-PAP - RH	1.08	1.03	1.1	1.03	100	100
GPO-PAP - AA	GPO-PAP - SD	1.08	1.02	1.1	1.03	100	100
GPO-PAP - BC AU	GPO-PAP - RCc	1.11	1.04	1.1	1.08	100	100
GPO-PAP - BC AU	GPO-PAP - RH	1.11	1.03	1.1	1.03	100	100
GPO-PAP - BC AU	GPO-PAP - SD	1.11	1.02	1.1	1.03	100	100
GPO-PAP - RCc	GPO-PAP - RH	1.04	1.03	1.08	1.03	100	100
GPO-PAP - RCc	GPO-PAP - SD	1.04	1.02	1.08	1.03	100	100
GPO-PAP - RH	GPO-PAP - SD	1.03	1.02	1.03	1.03	100	100
URATE							
Uricase - BC AU	Uricase,POD - AA	318	316.14	325.65	313.86	100	100
Uricase - BC AU	Uricase,POD - BC AU	318	320.43	325.65	327.42	100	100
Uricase - BC AU	Uricase,POD - RCc	318	311	325.65	309.83	100	100
Uricase - BC AU	Uricase,POD - RCI	318	322.22	325.65	317.67	100	100
Uricase - BC AU	Uricase,POD - RH	318	319.38	325.65	324.42	100	100
Uricase - BC AU	Uricase - RCc	318	319.17	325.65	318.17	100	100
Uricase - BC AU	Uricase - RCI	318	318	325.65	317.57	100	100
Uricase - BC AU	Uricase - RH	318	319	325.65	322.33	100	100
Uricase - BC AU	Uricase - SD	318	308.6	325.65	315.9	100	100
Uricase,POD - AA	Uricase,POD - BC AU	316.14	320.43	313.86	327.42	100	100
Uricase,POD - AA	Uricase,POD - RCc	316.14	311	313.86	309.83	100	100
Uricase,POD - AA	Uricase,POD - RCI	316.14	322.22	313.86	317.67	100	100
Uricase,POD - AA	Uricase,POD - RH	316.14	319.38	313.86	324.42	100	100

Uricase,POD - AA	Uricase - RCc	316.14	319.17	313.86	318.17	100	100
Uricase,POD - AA	Uricase - RCI	316.14	318	313.86	317.57	100	100
Uricase,POD - AA	Uricase - RH	316.14	319	313.86	322.33	100	100
Uricase,POD - AA	Uricase - SD	316.14	308.6	313.86	315.9	100	100
Uricase,POD - BC AU	Uricase,POD - RCc	320.43	311	327.42	309.83	100	100
Uricase,POD - BC AU	Uricase,POD - RCI	320.43	322.22	327.42	317.67	100	100
Uricase,POD - BC AU	Uricase,POD - RH	320.43	319.38	327.42	324.42	100	100
Uricase,POD - BC AU	Uricase - RCc	320.43	319.17	327.42	318.17	100	100
Uricase,POD - BC AU	Uricase - RCI	320.43	318	327.42	317.57	100	100
Uricase,POD - BC AU	Uricase - RH	320.43	319	327.42	322.33	100	100
Uricase,POD - BC AU	Uricase - SD	320.43	308.6	327.42	315.9	100	100
Uricase,POD - RCc	Uricase,POD - RCI	311	322.22	309.83	317.67	100	100
Uricase,POD - RCc	Uricase,POD - RH	311	319.38	309.83	324.42	100	100
Uricase,POD - RCc	Uricase - RCc	311	319.17	309.83	318.17	100	100
Uricase,POD - RCc	Uricase - RCI	311	318	309.83	317.57	100	100
Uricase,POD - RCc	Uricase - RH	311	319	309.83	322.33	100	100
Uricase,POD - RCc	Uricase - SD	311	308.6	309.83	315.9	100	100
Uricase,POD - RCI	Uricase,POD - RH	322.22	319.38	317.67	324.42	100	100
Uricase,POD - RCI	Uricase - RCc	322.22	319.17	317.67	318.17	100	100
Uricase,POD - RCI	Uricase - RCI	322.22	318	317.67	317.57	100	100
Uricase,POD - RCI	Uricase - RH	322.22	319	317.67	322.33	100	100
Uricase,POD - RCI	Uricase - SD	322.22	308.6	317.67	315.9	100	100
Uricase,POD - RH	Uricase - RCc	319.38	319.17	324.42	318.17	100	100
Uricase,POD - RH	Uricase - RCI	319.38	318	324.42	317.57	100	100
Uricase,POD - RH	Uricase - RH	319.38	319	324.42	322.33	100	100
Uricase,POD - RH	Uricase - SD	319.38	308.6	324.42	315.9	100	100
Uricase - RCc	Uricase - RCI	319.17	318	318.17	317.57	100	100
Uricase - RCc	Uricase - RH	319.17	319	318.17	322.33	100	100
Uricase - RCc	Uricase - SD	319.17	308.6	318.17	315.9	100	100
Uricase - RCI	Uricase - RH	318	319	317.57	322.33	100	100
Uricase - RCI	Uricase - SD	318	308.6	317.57	315.9	100	100

Uricase - RH	Uricase - SD	319	308.6	322.33	315.9	100	100
			UREA				
Urease,GLDH - AA	Urease,GLDH - BC AU	6.41	6.5	4.85	4.97	99.3	97.7
Urease,GLDH - AA	Urease,GLDH - HP	6.41	6.17	4.85	4.82	99.9	100
Urease,GLDH - AA	Urease,GLDH - RCc	6.41	6.32	4.85	4.81	100	100
Urease,GLDH - AA	Urease,GLDH - RCI	6.41	6.11	4.85	4.71	87.8	96.5
Urease,GLDH - AA	Urease,GLDH - RH	6.41	6.55	4.85	4.94	99.6	99.8
Urease,GLDH - AA	Urease,GLDH - SD	6.41	6.61	4.85	5.03	99.7	99.8
Urease,GLDH - BC AU	Urease,GLDH - HP	6.5	6.17	4.97	4.82	100	100
Urease,GLDH - BC AU	Urease,GLDH - RCc	6.5	6.32	4.97	4.81	100	100
Urease,GLDH - BC AU	Urease,GLDH - RCI	6.5	6.11	4.97	4.71	68.8	89.7
Urease,GLDH - BC AU	Urease,GLDH - RH	6.5	6.55	4.97	4.94	100	100
Urease,GLDH - BC AU	Urease,GLDH - SD	6.5	6.61	4.97	5.03	100	100
Urease,GLDH - HP	Urease,GLDH - RCc	6.17	6.32	4.82	4.81	100	100
Urease,GLDH - HP	Urease,GLDH - RCI	6.17	6.11	4.82	4.71	100	100
Urease,GLDH - HP	Urease,GLDH - RH	6.17	6.55	4.82	4.94	100	100
Urease,GLDH - HP	Urease,GLDH - SD	6.17	6.61	4.82	5.03	100	100
Urease,GLDH - RCc	Urease,GLDH - RCI	6.32	6.11	4.81	4.71	100	100
Urease,GLDH - RCc	Urease,GLDH - RH	6.32	6.55	4.81	4.94	100	100
Urease,GLDH - RCc	Urease,GLDH - SD	6.32	6.61	4.81	5.03	100	100
Urease,GLDH - RCI	Urease,GLDH - RH	6.11	6.55	4.71	4.94	97.4	99.5
Urease,GLDH - RCI	Urease,GLDH - SD	6.11	6.61	4.71	5.03	97.3	99.3
Urease,GLDH - RH	Urease,GLDH - SD	6.55	6.61	4.94	5.03	100	100

Commutability of control C1/2016 was evaluated for 426 MP combinations. The concentration ranges for most analytes correspond to normal or low pathological level according to appropriate reference intervals.

The control was found fully commutable for 8 analytes: AST, CK, iron, LDH, phosphate, sodium, triglycerides and urate. High commutability was also observed for AMY, glucose, potassium, bilirubin, proteins, and urea.

The control is almost fully noncommutable for HDL cholesterol, with results of 17/21 (81%) evaluated pairwise combinations of MPs showing less than 95% commutability. The number of MPs used for chloride determination was only three, out of which two combinations were found to be noncommutable with control sample C1/2016. Moderate commutability was observed for ALT, AP, calcium, cholesterol, creatinine and GGT.

For four analytes (ALT, cholesterol, chloride and GGT), instrument SD and the appropriate method was most likely the source of noncommutability, since control sample was found to be noncommutable for almost all the MPs combinations involving SD. Analysing patterns of noncommutability in other pairwise MP combinations, one can identify further possible sources of noncommutability among instruments: Arsenaso III-BC AU for calcium, compensated and noncompensated Jaffe methods from BC AU for creatinine and diazo-BC AU for bilirubin measurement. For HDL and AP, it is hard to identify such MPs because the control sample was found noncommutable with many MP combinations.

The MPs that show nonharmonisation on serum sample, very often show to be noncommutable in control sample (Table 15). Out of 31 nonharmonised MPs, 25 shows also noncommutability. If only harmonised MPs were assessed, the control C1/2016 would also be fully commutable for ALT. AMY and urea; the total of 11/22 analytes. Overall commutability would also improve for AP, calcium and creatinine.

Table 15. Contingency table showing the number of commutable/noncommutable and harmonised/nonharmonised MP combinations in the EQA survey 1**.**

	SURVEY 1/2016		
	C	**NC**	**Total**
H	348	47	395
NH	6	25	31
Total	354	72	426

H-harmonised, NH – nonharmonised, C – commutable, NC - noncommutable

4.2.2.2 Commutability evaluation of control sample C2/2016 using the false flagging method

Prior to evaluation of commutability of control C2/2016 used in the second EQA survey, the spiked serum sample used for comparison with control sample was also assessed for commutability to check whether this kind of a sample might be used as a substitute for native serum. The commutability of spiked serum sample was assessed using the false flagging method in comparison of results from spiked and serum sample prepared in the course of the third EQA survey. The spiked serum sample was initially considered the control sample and the MPs were expected to have the same flagging rate on these samples as in native serum. After performing the analysis, a total of 48 MP pairs were excluded across 12 analytes. The exclusion was solely based on commutability results, irrespective of concentration levels of each. Out of those 48 excluded MP pairs, 19 showed also nonharmonisation when evaluated on native samples only. Particularly high rate of excluded MP pairs was observed for creatinine (16/36) and bilirubin (9/21). It must be noticed that creatinine was not even the analyte used for spiking. Two MPs were also excluded for chloride, the analyte used for spiking, which resulted in only one MPs combination to be further evaluated. The results of all excluded MPs in the second EQA survey are shown in Appendix. Table 16 presents the results of harmonisation and commutability for evaluated MP combinations in EQA survey 2 for 22 assessed analytes. Percentage harmonisation and percentage commutability below 95 indicate nonharmonisation and noncommutability for named MP pair.

Table 16. The results of commutability evaluation of EQA control sample C2/2016 using the false flagging method.

		EQA SURVEY 2					
MP 1	**MP 2**	**Mean MP1 (serum)**	**Mean MP2 (serum)**	**Mean MP1 (control)**	**Mean MP2 (control)**	**% harmo-nisation**	**% commu-tability**
			ALT				
IFCC- BC AU	Photometry UV- AA	25.53	24	152.21	149.5	99.9	99.9
IFCC- BC AU	Photometry UV- BC AU	25.53	24.35	152.21	151.94	100	100
IFCC- BC AU	Photometry UV- RCc	25.53	23.25	152.21	146.89	100	100
IFCC- BC AU	Photometry UV- RCI	25.53	23	152.21	141.07	100	100

IFCC- BC AU	Photometry UV- RH	25.53	23.42	152.21	146.1	100	100
Photometry UV- AA	Photometry UV- BC AU	24	24.35	149.5	151.94	100	100
Photometry UV- AA	Photometry UV- RCc	24	23.25	149.5	146.89	100	100
Photometry UV- AA	Photometry UV- RCI	24	23	149.5	141.07	100	100
Photometry UV- AA	Photometry UV- RH	24	23.42	149.5	146.1	100	100
Photometry UV- BC AU	Photometry UV- RCc	24.35	23.25	151.94	146.89	100	100
Photometry UV- BC AU	Photometry UV- RCI	24.35	23	151.94	141.07	100	100
Photometry UV- BC AU	Photometry UV- RH	24.35	23.42	151.94	146.1	100	100
Photometry UV- RCc	Photometry UV- RCI	23.25	23	146.89	141.07	100	100
Photometry UV- RCc	Photometry UV- RH	23.25	23.42	146.89	146.1	100	100
Photometry UV- RCI	Photometry UV- RH	23	23.42	141.07	146.1	100	100
			ALP				
IFCC- AA	IFCC- BC AU	64.18	69.44	288.64	311.8	95.6	59.7
IFCC- AA	IFCC- RCc	64.18	64.85	288.64	267.38	100	91.1
IFCC- AA	IFCC- RCI	64.18	67.47	288.64	272.39	100	96.7
IFCC- AA	IFCC- RCMira	64.18	70.5	288.64	288	100	97.3
IFCC- AA	IFCC- RH	64.18	68.68	288.64	291.16	99.7	99.6
IFCC- AA	IFCC- SD	64.18	67.83	288.64	283.64	100	100
IFCC- BC AU	IFCC- RCc	69.44	64.85	311.8	267.38	100	0
IFCC- BC AU	IFCC- RCI	69.44	67.47	311.8	272.39	100	4.9
IFCC- BC AU	IFCC- RCMira	69.44	70.5	311.8	288	100	29.8
IFCC- BC AU	IFCC- RH	69.44	68.68	311.8	291.16	100	76.7
IFCC- BC AU	IFCC- SD	69.44	67.83	311.8	283.64	100	80.1
IFCC- RCc	IFCC- RCI	64.85	67.47	267.38	272.39	100	100
IFCC- RCc	IFCC- RCMira	64.85	70.5	267.38	288	100	59.3
IFCC- RCc	IFCC- RH	64.85	68.68	267.38	291.16	100	99.8
IFCC- RCc	IFCC- SD	64.85	67.83	267.38	283.64	100	100
IFCC- RCI	IFCC- RCMira	67.47	70.5	272.39	288	100	63.8
IFCC- RCI	IFCC- RH	67.47	68.68	272.39	291.16	100	100
IFCC- RCI	IFCC- SD	67.47	67.83	272.39	283.64	100	100
IFCC- RCMira	IFCC- RH	70.5	68.68	288	291.16	100	90
IFCC- RCMira	IFCC- SD	70.5	67.83	288	283.64	100	96
IFCC- RH	IFCC- SD	68.68	67.83	291.16	283.64	100	100
			AMY				
IFCC- AA	IFCC- BC AU	51	52.83	355.3	344.3	100	100
IFCC- AA	IFCC- RCc	51	51.85	355.3	330.69	100	100
IFCC- AA	IFCC- RCI	51	53.08	355.3	331.73	100	100
IFCC- AA	IFCC- RH	51	49.68	355.3	324.65	100	100
IFCC- AA	CNP-G3- SD	51	47.71	355.3	357.29	100	100
IFCC- BC AU	IFCC- RCc	52.83	51.85	344.3	330.69	100	100
IFCC- BC AU	IFCC- RCI	52.83	53.08	344.3	331.73	100	100

IFCC- BC AU	CNP-G3- SD	52.83	47.71	344.3	357.29	83.1	83.5
IFCC- RCc	IFCC- RCI	51.85	53.08	330.69	331.73	100	100
IFCC- RCc	IFCC- RH	51.85	49.68	330.69	324.65	100	100
IFCC- RCc	CNP-G3- SD	51.85	47.71	330.69	357.29	99.9	99.9
IFCC- RCI	IFCC- RH	53.08	49.68	331.73	324.65	100	100
IFCC- RCI	CNP-G3- SD	53.08	47.71	331.73	357.29	100	100
IFCC- RH	CNP-G3- SD	49.68	47.71	324.65	357.29	100	99.9
AST							
IFCC- BC AU	IFCC- SD	23.33	23.5	229.37	236.42	100	100
IFCC- BC AU	Photometry UV- AA	23.33	19.9	229.37	208.7	75.5	76.5
IFCC- BC AU	Photometry UV- BC AU	23.33	22.65	229.37	226.31	100	100
IFCC- BC AU	Photometry UV- RCc	23.33	20.12	229.37	205.56	99.6	99.7
IFCC- BC AU	Photometry UV- RCI	23.33	20.71	229.37	206.73	100	100
IFCC- BC AU	Photometry UV- RH	23.33	20.74	229.37	205	100	100
IFCC- SD	Photometry UV- AA	23.5	19.9	236.42	208.7	99.2	99.2
IFCC- SD	Photometry UV- BC AU	23.5	22.65	236.42	226.31	100	100
IFCC- SD	Photometry UV- RCc	23.5	20.12	236.42	205.56	100	100
IFCC- SD	Photometry UV- RCI	23.5	20.71	236.42	206.73	100	100
IFCC- SD	Photometry UV- RH	23.5	20.74	236.42	205	98.4	98.5
Photometry UV- AA	Photometry UV- BC AU	19.9	22.65	208.7	226.31	94.5	94.5
Photometry UV- AA	Photometry UV- RCc	19.9	20.12	208.7	205.56	100	100
Photometry UV- AA	Photometry UV- RCI	19.9	20.71	208.7	206.73	100	100
Photometry UV- AA	Photometry UV- RH	19.9	20.74	208.7	205	100	100
Photometry UV- BC AU	Photometry UV- RCc	22.65	20.12	226.31	205.56	100	100
Photometry UV- BC AU	Photometry UV- RCI	22.65	20.71	226.31	206.73	100	100
Photometry UV- BC AU	Photometry UV- RH	22.65	20.74	226.31	205	100	100
Photometry UV- RCc	Photometry UV- RCI	20.12	20.71	205.56	206.73	100	100
Photometry UV- RCc	Photometry UV- RH	20.12	20.74	205.56	205	100	100
Photometry UV- RCI	Photometry UV- RH	20.71	20.74	206.73	205	100	100
CALCIUM							
Asenaso III- AA	Asenaso III- BC AU	2.18	2.25	3.08	3.17	99.3	74.4
Asenaso III- AA	NM-BAPTA- RCI	2.18	2.23	3.08	3.18	99.9	98.9
Asenaso III- AA	cresolphthalein - BC AU	2.18	2.22	3.08	3.16	100	99.5

Asenaso III-BC AU	NM-BAPTA-RC	2.25	2.23	3.17	3.18	100	100
NM-BAPTA-RC	cresolphthalein - BC AU	2.23	2.22	3.18	3.16	100	100
NM-BAPTA-RC	cresolphthalein - SD	2.23	2.19	3.18	3.08	100	93.4
cresolphthalein - BC AU	cresolphthalein - SD	2.22	2.19	3.16	3.08	100	96.5
CHLORIDE							
Indirect ISE-AA	Indirect ISE-SD	122.08	122.71	135.42	130.14	98.8	80.9
CHOLESTEROL							
CHOD-PAP-AA	CHOD-PAP-BC AU	5.27	5.44	6.05	6.23	100	100
CHOD-PAP-AA	CHOD-PAP-RCc	5.27	5.17	6.05	5.94	100	100
CHOD-PAP-AA	CHOD-PAP-RCI	5.27	5.28	6.05	6.06	100	100
CHOD-PAP-AA	CHOD-PAP-RH	5.27	5.26	6.05	6.04	100	100
CHOD-PAP-AA	CHOD-PAP-SD	5.27	5.24	6.05	5.42	100	10.5
CHOD-PAP-BC AU	CHOD-PAP-RCc	5.44	5.17	6.23	5.94	100	100
CHOD-PAP-BC AU	CHOD-PAP-RCI	5.44	5.28	6.23	6.06	100	100
CHOD-PAP-BC AU	CHOD-PAP-RH	5.44	5.26	6.23	6.04	100	100
CHOD-PAP-BC AU	CHOD-PAP-SD	5.44	5.24	6.23	5.42	100	0
CHOD-PAP-RCc	CHOD-PAP-RCI	5.17	5.28	5.94	6.06	100	100
CHOD-PAP-RCc	CHOD-PAP-RH	5.17	5.26	5.94	6.04	100	100
CHOD-PAP-RCc	CHOD-PAP-SD	5.17	5.24	5.94	5.42	100	21.7
CHOD-PAP-RCI	CHOD-PAP-RH	5.28	5.26	6.06	6.04	100	100
CHOD-PAP-RCI	CHOD-PAP-SD	5.28	5.24	6.06	5.42	100	1
CHOD-PAP-RH	CHOD-PAP-SD	5.26	5.24	6.04	5.42	100	1.9
CK							
IFCC- AA	IFCC- BC AU	150.73	155.94	240.64	248.85	100	100
IFCC- AA	IFCC- RCc	150.73	153.67	240.64	251.29	100	99.7
IFCC- AA	IFCC- RCI	150.73	155.82	240.64	247.92	100	100
IFCC- AA	IFCC- RH	150.73	151.3	240.64	242.7	100	100
IFCC- AA	IFCC- SD	150.73	150.43	240.64	237.57	100	100
IFCC- BC AU	IFCC- RCc	155.94	153.67	248.85	251.29	100	100
IFCC- BC AU	IFCC- RCI	155.94	155.82	248.85	247.92	100	100
IFCC- BC AU	IFCC- RH	155.94	151.3	248.85	242.7	100	100
IFCC- BC AU	IFCC- SD	155.94	150.43	248.85	237.57	100	100
IFCC- RCc	IFCC- RCI	153.67	155.82	251.29	247.92	100	100
IFCC- RCc	IFCC- RH	153.67	151.3	251.29	242.7	100	100
IFCC- RCc	IFCC- SD	153.67	150.43	251.29	237.57	100	100

IFCC- RCI	IFCC- RH	155.82	151.3	247.92	242.7	100	100
IFCC- RCI	IFCC- SD	155.82	150.43	247.92	237.57	100	100
IFCC- RH	IFCC- SD	151.3	150.43	242.7	237.57	100	100
CREATININE							
Compensated Jaffe- AA	Compensated Jaffe- SD	77.29	82.44	273.67	252.56	100	99.6
Compensated Jaffe- AA	Non-compensated Jaffe- BC AU	77.29	87.77	273.67	240.87	3.2	42.6
Compensated Jaffe- BC AU	Enzymatic method- BC AU	76.24	77.58	237.4	249.55	100	100
Compensated Jaffe- RCc	Compensated Jaffe- RCI	82.82	80	257.64	241	97.1	96.7
Compensated Jaffe- RCc	Compensated Jaffe- RH	82.82	83.41	257.64	252.06	99.9	99.9
Compensated Jaffe- RCc	Compensated Jaffe- SD	82.82	82.44	257.64	252.56	100	100
Compensated Jaffe- RCc	Enzymatic method- BC AU	82.82	77.58	257.64	249.55	87	96.4
Compensated Jaffe- RCI	Compensated Jaffe- RH	80	83.41	241	252.06	97	99.9
Compensated Jaffe- RCI	Compensated Jaffe- SD	80	82.44	241	252.56	100	100
Compensated Jaffe- RCI	Enzymatic method- BC AU	80	77.58	241	249.55	100	100
Compensated Jaffe- RH	Compensated Jaffe- SD	83.41	82.44	252.06	252.56	100	100
Compensated Jaffe- SD	Enzymatic method- BC AU	82.44	77.58	252.56	249.55	100	100
GGT							
IFCC- AA	IFCC- BC AU	27.17	26.26	144.75	144.78	100	100
IFCC- AA	IFCC- RCc	27.17	25.91	144.75	147	100	100
IFCC- AA	IFCC- RCI	27.17	25.83	144.75	150	100	100
IFCC- AA	IFCC- RH	27.17	26.92	144.75	147.36	100	100
IFCC- AA	IFCC- SD	27.17	28.38	144.75	147.92	100	100
IFCC- BC AU	IFCC- RCc	26.26	25.91	144.78	147	100	100
IFCC- BC AU	IFCC- RCI	26.26	25.83	144.78	150	100	100
IFCC- BC AU	IFCC- RH	26.26	26.92	144.78	147.36	100	100
IFCC- BC AU	IFCC- SD	26.26	28.38	144.78	147.92	90.7	90.8
IFCC- RCc	IFCC- RCI	25.91	25.83	147	150	100	100
IFCC- RCc	IFCC- RH	25.91	26.92	147	147.36	100	100
IFCC- RCc	IFCC- SD	25.91	28.38	147	147.92	100	100
IFCC- RCI	IFCC- RH	25.83	26.92	150	147.36	100	100
IFCC- RCI	IFCC- SD	25.83	28.38	150	147.92	98.9	98.9
IFCC- RH	IFCC- SD	26.92	28.38	147.36	147.92	100	100
GLUCOSE							
GOD-PAP- BC AU	GOD-PAP- RH	10.13	9.9	10.42	10.31	100	100
GOD-PAP- BC AU	Hexokinase- AA	10.13	9.94	10.42	10.41	100	100

GOD-PAP- BC AU	Hexokinase- BC AU	10.13	10.14	10.42	10.56	100	99.9
GOD-PAP- BC AU	Hexokinase- RCc	10.13	9.94	10.42	10.42	100	100
GOD-PAP- BC AU	Hexokinase- RCI	10.13	10.08	10.42	10.44	100	100
GOD-PAP- BC AU	Hexokinase- SD	10.13	9.95	10.42	10.65	100	99.8
GOD-PAP- RH	Hexokinase- AA	9.9	9.94	10.31	10.41	100	100
GOD-PAP- RH	Hexokinase- BC AU	9.9	10.14	10.31	10.56	100	100
GOD-PAP- RH	Hexokinase- RCc	9.9	9.94	10.31	10.42	100	100
GOD-PAP- RH	Hexokinase- RCI	9.9	10.08	10.31	10.44	100	100
GOD-PAP- RH	Hexokinase- SD	9.9	9.95	10.31	10.65	100	100
Hexokinase- AA	Hexokinase- BC AU	9.94	10.14	10.41	10.56	100	100
Hexokinase- AA	Hexokinase- RCc	9.94	9.94	10.41	10.42	100	100
Hexokinase- AA	Hexokinase- RCI	9.94	10.08	10.41	10.44	100	100
Hexokinase- AA	Hexokinase- SD	9.94	9.95	10.41	10.65	100	100
Hexokinase- BC AU	Hexokinase- RCc	10.14	9.94	10.56	10.42	100	100
Hexokinase- BC AU	Hexokinase- RCI	10.14	10.08	10.56	10.44	100	100
Hexokinase- BC AU	Hexokinase- SD	10.14	9.95	10.56	10.65	99.1	99.6
Hexokinase- RCc	Hexokinase- RCI	9.94	10.08	10.42	10.44	100	100
Hexokinase- RCc	Hexokinase- SD	9.94	9.95	10.42	10.65	100	100
Hexokinase- RCI	Hexokinase- SD	10.08	9.95	10.44	10.65	100	100
			HDL				
Homogenous- AA	Homogenous- BC AU	1.21	1.19	1.92	1.61	100	0.1
Homogenous- AA	Homogenous- RCc	1.21	1.18	1.92	1.41	100	0
Homogenous- AA	Homogenous- RCI	1.21	1.19	1.92	1.45	100	0
Homogenous- AA	Homogenous- RH	1.21	1.28	1.92	1.81	99.9	84.1
Homogenous- AA	Homogenous- SD	1.21	1.25	1.92	1.47	99.9	0
Homogenous- BC AU	Homogenous- RCc	1.19	1.18	1.61	1.41	100	1
Homogenous- BC AU	Homogenous- RCI	1.19	1.19	1.61	1.45	100	36.3
Homogenous- BC AU	Homogenous- RH	1.19	1.28	1.61	1.81	58	59
Homogenous- RCc	Homogenous- RCI	1.18	1.19	1.41	1.45	100	99.7

Homogenous-RCc	Homogenous-RH	1.18	1.28	1.41	1.81	97.9	0
Homogenous-RCc	Homogenous-SD	1.18	1.25	1.41	1.47	97.5	98.6
Homogenous-RCI	Homogenous-RH	1.19	1.28	1.45	1.81	98.9	4.9
Homogenous-RCI	Homogenous-SD	1.19	1.25	1.45	1.47	97.5	99.4
Homogenous-RH	Homogenous-SD	1.28	1.25	1.81	1.47	99.9	0.7
			IRON				
Ferene- AA	Ferene- HP	11.7	12.67	31.72	31.57	99.5	99.5
Ferene- AA	Ferene- RH	11.7	12.49	31.72	33.17	100	100
Ferene- AA	Ferene- SD	11.7	12.83	31.72	30.81	88.5	88.5
Ferene- AA	Ferrozine- RCc	11.7	13.34	31.72	33.31	94.3	94.3
Ferene- AA	Ferrozine- RCI	11.7	13.13	31.72	33.35	95.1	95.1
Ferene- AA	Ferrozine- RH	11.7	12.92	31.72	32.67	100	100
Ferene- AA	TPTZ- BC AU	11.7	12.31	31.72	32.23	100	100
Ferene- HP	Ferene- RH	12.67	12.5	31.5	32.58	100	100
Ferene- HP	Ferene- SD	12.67	12.83	31.57	30.81	97.2	97.2
Ferene- HP	Ferrozine- RCc	12.67	13.38	31.5	33	100	100
Ferene- HP	Ferrozine- RCI	12.67	13.19	31.5	33.53	100	100
Ferene- HP	Ferrozine- RH	12.67	13	31.5	32.67	100	100
Ferene- HP	TPTZ- BC AU	12.67	12.25	31.5	32.25	100	100
Ferene- RH	Ferene- SD	12.49	12.83	33.17	30.81	98.5	98.6
Ferene- RH	Ferrozine- RCc	12.5	13.38	32.58	33	100	100
Ferene- RH	Ferrozine- RCI	12.5	13.19	32.58	33.53	100	100
Ferene- RH	Ferrozine- RH	12.5	13	32.58	32.67	100	100
Ferene - RH	TPTZ- BC AU	12.5	12.25	32.58	32.25	100	100
Ferene- SD	Ferrozine- RCc	12.83	13.34	30.81	33.31	100	100
Ferene- SD	Ferrozine- RCI	12.83	13.13	30.81	33.35	98.8	98.8
Ferene- SD	Ferrozine- RH	12.83	12.92	30.81	32.67	99.7	99.7
Ferene- SD	TPTZ- BC AU	12.83	12.31	30.81	32.23	76.4	76.5
Ferrozine- RCc	Ferrozine- RCI	13.38	13.19	33	33.53	100	100
Ferrozine- RCc	Ferrozine- RH	13.38	13	33	32.67	100	100
Ferrozine- RCc	TPTZ- BC AU	13.38	12.25	33	32.25	100	100
Ferrozine- RCI	Ferrozine- RH	13.19	13	33.53	32.67	100	100
Ferrozine- RCI	TPTZ- BC AU	13.19	12.25	33.53	32.25	100	100
Ferrozine- RH	TPTZ- BC AU	13	12.25	32.67	32.25	100	100
			LDH				
IFCC- AA	IFCC- BC AU	144.6	147.61	312.64	316.75	100	100
IFCC- AA	IFCC- RCc	144.6	144.88	312.64	308.25	100	100
IFCC- AA	IFCC- RCI	144.6	149.46	312.64	317.29	100	100
IFCC- AA	IFCC- RH	144.6	141.38	312.64	308.62	100	100
IFCC- AA	IFCC- SD	144.6	143.17	312.64	312.5	100	100
IFCC- BC AU	IFCC- RCc	147.61	144.88	316.75	308.25	100	100
IFCC- BC AU	IFCC- RCI	147.61	149.46	316.75	317.29	100	100
IFCC- BC AU	IFCC- RH	147.61	141.38	316.75	308.62	100	100
IFCC- RCc	IFCC- RCI	144.88	149.46	308.25	317.29	100	100
IFCC- RCc	IFCC- RH	144.88	141.38	308.25	308.62	100	100
IFCC- RCc	IFCC- SD	144.88	143.17	308.25	312.5	100	100
IFCC- RCI	IFCC- RH	149.46	141.38	317.29	308.62	100	100

IFCC- RH	IFCC- SD	141.38	143.17	308.62	312.5	100	100
PHOSPHATE							
Ammonium-molybdate- AA	Ammonium-molybdate- RCc	0.96	0.94	2.86	2.82	100	100
Ammonium-molybdate- AA	Ammonium-molybdate- RCI	0.96	0.99	2.86	2.99	100	100
Ammonium-molybdate- BC AU	Ammonium-molybdate- RCc	0.96	0.94	2.9	2.82	100	100
Ammonium-molybdate- BC AU	Ammonium-molybdate- RCI	0.96	0.99	2.9	2.99	100	100
Ammonium-molybdate- RCc	Ammonium-molybdate- RCI	0.94	0.99	2.82	2.99	100	100
POTASSIUM							
FES- CC	Indirect ISE- AA	4.12	4.06	5.78	5.82	100	100
FES- CC	Indirect ISE- BC AU	4.12	4.07	5.78	5.76	100	100
FES- CC	Indirect ISE- RCc	4.12	4.1	5.78	5.9	100	100
FES- CC	Indirect ISE- RCI	4.12	4.09	5.78	5.85	100	100
FES- CC	Indirect ISE- SD	4.12	4.04	5.78	5.82	100	100
Indirect ISE- AA	Indirect ISE- BC AU	4.06	4.07	5.82	5.76	100	100
Indirect ISE- AA	Indirect ISE- RCc	4.06	4.1	5.82	5.9	100	100
Indirect ISE- AA	Indirect ISE- RCI	4.06	4.09	5.81	5.85	100	100
Indirect ISE- AA	Indirect ISE- SD	4.06	4.04	5.82	5.82	100	100
Indirect ISE- BC AU	Indirect ISE- RCc	4.07	4.1	5.76	5.9	100	100
Indirect ISE- BC AU	Indirect ISE- RCI	4.07	4.09	5.76	5.85	100	100
Indirect ISE- BC AU	Indirect ISE- SD	4.07	4.04	5.76	5.82	100	100
Indirect ISE- RCc	Indirect ISE- RCI	4.1	4.09	5.9	5.85	100	100
Indirect ISE- RCc	Indirect ISE- SD	4.1	4.04	5.9	5.82	100	100
Indirect ISE- RCI	Indirect ISE- SD	4.09	4.04	5.85	5.82	100	100
SODIUM							
FES- CC	Indirect ISE- AA	153.68	153.85	163.64	165	100	100
FES- CC	Indirect ISE- BC AU	153.68	153.56	163.64	164.01	100	100
FES- CC	Indirect ISE- RCc	153.68	153.91	163.64	166.82	100	98.2
FES- CC	Indirect ISE- RCI	153.68	154.18	163.64	166.45	100	100

FES- CC	Indirect ISE-SD	153.68	154.82	163.64	164.6	100	100
Indirect ISE-AA	Indirect ISE-BC AU	153.85	153.56	165	164.01	100	100
Indirect ISE-AA	Indirect ISE-RCc	153.85	153.91	165	166.82	100	100
Indirect ISE-AA	Indirect ISE-RCI	153.85	154.18	165	166.45	100	100
Indirect ISE-AA	Indirect ISE-SD	153.85	154.82	165	164.6	100	100
Indirect ISE-BC AU	Indirect ISE-RCc	153.56	153.91	164.01	166.82	100	89.7
Indirect ISE-BC AU	Indirect ISE-RCI	153.56	154.18	164.01	166.45	100	100
Indirect ISE-BC AU	Indirect ISE-SD	153.56	154.82	164.01	164.6	100	100
Indirect ISE-RCc	Indirect ISE-RCI	153.91	154.18	166.82	166.45	100	100
Indirect ISE-RCc	Indirect ISE-SD	153.91	154.82	166.82	164.6	100	100
Indirect ISE-RCI	Indirect ISE-SD	154.18	154.82	166.45	164.6	100	100
BILIRUBIN							
Diazo- AA	Diazo- RCc	31.7	27.38	88.6	80.31	24.6	63.9
Diazo- AA	Diazo- RCI	31.7	27.56	88.6	78.31	23.1	80.7
Diazo- BC AU	Diazo- HP	32.28	30.5	81.49	83.17	100	100
Diazo- BC AU	Diazo- RH	32.28	29.52	81.49	78.92	100	100
Diazo- HP	Diazo- RH	30.5	29.52	83.17	78.92	100	100
Diazo- HP	Diazo- SD	30.5	29.15	83.17	81.69	100	100
Diazo- RCc	Diazo- RCI	27.38	27.56	80.31	78.31	100	100
Diazo- RCc	Diazo- RH	27.38	29.52	80.31	78.92	96.7	98
Diazo- RCc	Diazo- SD	27.38	29.15	80.31	81.69	99.8	99.8
Diazo- RCI	Diazo- RH	27.56	29.52	78.31	78.92	100	100
Diazo- RCI	Diazo- SD	27.56	29.15	78.31	81.69	100	100
Diazo- RH	Diazo- SD	29.52	29.15	78.92	81.69	100	100
PROTEINS							
Biuret- AA	Biuret- RCc	64.9	66.64	93.7	93.14	100	100
Biuret- AA	Biuret- RCI	64.9	65.67	93.7	92.33	100	100
Biuret- AA	Biuret- RH	64.9	65.78	93.7	93	100	100
Biuret- AA	Biuret- SD	64.9	68.75	93.7	96.57	100	100
Biuret- BC AU	Biuret- RCc	66.45	66.64	94.08	93.14	100	100
Biuret- BC AU	Biuret- RCI	66.45	65.67	94.08	92.33	94.5	98.6
Biuret- BC AU	Biuret- RH	66.45	65.78	94.08	93	100	100
Biuret- BC AU	Biuret- SD	66.45	68.75	94.08	96.57	100	100
Biuret- RCc	Biuret- RCI	66.64	65.67	93.14	92.33	99.5	99.9
Biuret- RCc	Biuret- RH	66.64	65.78	93.14	93	100	100
Biuret- RCc	Biuret- SD	66.64	68.75	93.14	96.57	100	100
Biuret- RCI	Biuret- RH	65.67	65.78	92.33	93	99.8	99.9
Biuret- RCI	Biuret- SD	65.67	68.75	92.33	96.57	96.5	98.7
Biuret- RH	Biuret- SD	65.78	68.75	93	96.57	99.9	99.9
TRIGLYCERIDES							
GPO-PAP- AA	GPO-PAP- BC AU	1.69	1.76	4.57	4.63	100	100

GPO-PAP- AA	GPO-PAP-RCc	1.69	1.64	4.57	4.22	100	100
GPO-PAP- AA	GPO-PAP-RCI	1.69	1.64	4.57	4.45	100	100
GPO-PAP- AA	GPO-PAP- RH	1.69	1.66	4.57	4.37	100	100
GPO-PAP- AA	GPO-PAP- SD	1.69	1.7	4.57	4.51	100	100
GPO-PAP- BC AU	GPO-PAP-RCc	1.76	1.65	4.63	4.23	100	100
GPO-PAP- BC AU	GPO-PAP-RCI	1.76	1.68	4.63	4.45	100	100
GPO-PAP- BC AU	GPO-PAP- RH	1.76	1.65	4.63	4.37	100	100
GPO-PAP- BC AU	GPO-PAP- SD	1.76	1.7	4.63	4.52	100	100
GPO-PAP-RCc	GPO-PAP-RCI	1.65	1.68	4.23	4.45	100	100
GPO-PAP-RCc	GPO-PAP- RH	1.65	1.65	4.23	4.37	100	100
GPO-PAP-RCc	GPO-PAP- SD	1.65	1.7	4.23	4.52	100	100
GPO-PAP-RCI	GPO-PAP- RH	1.68	1.65	4.45	4.37	100	100
GPO-PAP-RCI	GPO-PAP- SD	1.68	1.7	4.45	4.52	100	100
GPO-PAP- RH	GPO-PAP- SD	1.65	1.7	4.37	4.52	100	100
			URATE				
Uricase- BC AU	Uricase,POD-AA	341.96	339.14	688	690.71	100	99.8
Uricase- BC AU	Uricase,POD-BC AU	341.96	342.27	688	693.59	100	100
Uricase- BC AU	Uricase,POD-RCc	341.96	325	688	671.57	100	100
Uricase- BC AU	Uricase,POD-RCI	341.96	338.3	688	689.2	100	100
Uricase- BC AU	Uricase,POD-RH	341.96	345	688	696.38	100	100
Uricase- BC AU	Uricase- RH	341.96	342.08	688	697.08	100	100
Uricase- BC AU	Uricase- SD	341.96	315.5	688	679	100	100
Uricase,POD-AA	Uricase,POD-BC AU	339.14	342.27	690.71	693.59	99.9	100
Uricase,POD-AA	Uricase,POD-RCc	339.14	325	690.71	671.57	99.9	100
Uricase,POD-AA	Uricase,POD-RCI	339.14	338.3	690.71	689.2	100	100
Uricase,POD-AA	Uricase,POD-RH	339.14	345	690.71	696.38	100	100
Uricase,POD-AA	Uricase- RH	339.14	342.08	690.71	697.08	100	100
Uricase,POD-BC AU	Uricase,POD-RCc	342.27	325	693.59	671.57	100	100
Uricase,POD-BC AU	Uricase,POD-RCI	342.27	338.3	693.59	689.2	100	100
Uricase,POD-BC AU	Uricase,POD-RH	342.27	345	693.59	696.38	100	100
Uricase,POD-BC AU	Uricase- RH	342.27	342.08	693.59	697.08	100	100

Uricase,POD-RCc	Uricase,POD-RCI	325	338.3	671.57	689.2	100	100
Uricase,POD-RCc	Uricase,POD-RH	325	345	671.57	696.38	100	100
Uricase,POD-RCc	Uricase- RH	325	342.08	671.57	697.08	100	100
Uricase,POD-RCc	Uricase- SD	325	315.5	671.57	679	100	100
Uricase,POD-RCI	Uricase,POD-RH	338.3	345	689.2	696.38	100	100
Uricase,POD-RCI	Uricase- RH	338.3	342.08	689.2	697.08	100	100
Uricase,POD-RCI	Uricase - SD	338.3	315.5	689.2	679	100	100
Uricase,POD-RH	Uricase- RH	345	342.08	696.38	697.08	100	100
Uricase,POD-RH	Uricase- SD	345	315.5	696.38	679	100	100
Uricase- RH	Uricase- SD	342.08	315.5	697.08	679	100	100
			UREA				
Urease,GLDH-AA	Urease,GLDH-BC AU	9.38	9.58	12.89	13.02	100	100
Urease,GLDH-AA	Urease,GLDH-HP	9.38	9.27	12.89	12.78	100	100
Urease,GLDH-AA	Urease,GLDH-RCc	9.38	9.38	12.89	12.94	100	100
Urease,GLDH-AA	Urease,GLDH-RCI	9.38	9.38	12.89	12.86	100	100
Urease,GLDH-AA	Urease,GLDH-RH	9.38	9.56	12.89	13.03	100	100
Urease,GLDH-AA	Urease,GLDH-SD	9.38	9.54	12.89	13.12	100	100
Urease,GLDH-BC AU	Urease,GLDH-HP	9.58	9.27	13.02	12.78	100	100
Urease,GLDH-BC AU	Urease,GLDH-RCc	9.58	9.38	13.02	12.94	100	100
Urease,GLDH-BC AU	Urease,GLDH-RH	9.58	9.56	13.02	13.03	100	100
Urease,GLDH-HP	Urease,GLDH-RCc	9.27	9.38	12.78	12.94	100	100
Urease,GLDH-HP	Urease,GLDH-RCI	9.27	9.38	12.78	12.86	100	100
Urease,GLDH-HP	Urease,GLDH-RH	9.27	9.56	12.78	13.03	100	100
Urease,GLDH-HP	Urease,GLDH-SD	9.27	9.54	12.78	13.12	100	100
Urease,GLDH-RCc	Urease,GLDH-RCI	9.38	9.38	12.94	12.86	100	100
Urease,GLDH-RCc	Urease,GLDH-RH	9.38	9.56	12.94	13.03	100	100
Urease,GLDH-RCc	Urease,GLDH-SD	9.38	9.54	12.94	13.12	100	100
Urease,GLDH-RH	Urease,GLDH-SD	9.56	9.54	13.03	13.12	100	100

The concentration ranges assessed in the second EQA survey correspond mostly to high levels according to the reference intervals of each analyte. The concentration levels of serum samples to which the results of control samples were compared were in the normal ranges, except for analytes used for spiking: glucose, urea, sodium, chloride and bilirubin.

The control was evaluated for commutability with 331 MPs combinations. Full commutability was found for 11 analytes: ALT, CK, glucose, LDH, phosphate, potassium, proteins, triglycerides, urate and urea. Since the control also showed high commutability for AMY, AST, GGT, iron and sodium, one can conclude that patterns of commutability are the same for those analytes on lyophilised control samples from the same manufacturer at different concentration levels. The difference in analyte-related commutability of this control sample can be seen for GGT, having more commutable decisions at this high concentration level. Creatinine and bilirubin commutability conclusions for C2/2016 are somewhat different than in C1/2016, although many MP combinations were excluded from assessment in the second EQA survey to be able to compare the commutability of these controls for used MPs. Similar to C1/2016, this control also showed almost complete noncommutability for HDL cholesterol and rather low commutability for AP. One evaluated MP pair for chloride was also noncommutable.

Since SD was excluded from assessment for ALT, this instrument was not the cause for noncommutability of the MP combinations involving SD. In the case of cholesterol, all pairwise MP combinations with SD remained to be the probable source of noncommutability. As a difference from the EQA survey 1, the SD showed noncommutability in only one MP pair used for GGT measurement for the control sample C2/2016.

If commutability of C2/2016 is analysed in relation to the harmonisation of MPs, the overall commutability would be much better for AMY, AST, creatinine, GGT, iron and bilirubin, where full commutability of this control would be observed. The contingency table showing the relationship between harmonisation and commutability within this EQA survey is shown in Table 17.

Table 17. The contingency table showing the number of commutable/noncommutable and harmonised/nonharmonised MP combinations in the EQA survey 2**.**

	SURVEY 2/2016		
	C	**NC**	**Total**
H	288	29	317
NH	2	12	14
Total	290	41	331

H-harmonised, NH – nonharmonised, C – commutable, NC - noncommutable

4.2.2.3 Commutability evaluation of control sample C3/2016 using the false flagging method

The results of commutability evaluation of the EQA control sample C3/2016 is presented in Table 18. Mean values for each MP and sample evaluated in EQA survey 3 are presented in Table 18 together with the results for harmonisation and commutability of each pair of MPs. The MP pairs are considered harmonised/commutable when percentage harmonisation/commutability is ≥ 95%, as defined in Materials and methods section 3.5.

Table 18. The results of commutability evaluation of EQA control sample C3/2016 using the false flagging method

		EQA SURVEY 3					
MP 1	**MP 2**	**Mean MP1 (serum)**	**Mean MP2 (serum)**	**Mean MP1 (control)**	**Mean MP2 (control)**	**% harmo-nisation**	**% commu-tability**
			ALT				
IFCC- BC AU	IFCC- SD	7.27	13.25	32.76	33	0	0
IFCC- BC AU	Photometry UV- AA	7.27	7.1	32.76	32.2	100	100
IFCC- BC AU	Photometry UV- BC AU	7.27	6.63	32.76	32.82	100	100
IFCC- BC AU	Photometry UV- RCc	7.27	6.62	32.76	31	100	100
IFCC- BC AU	Photometry UV- RCI	7.27	6.4	32.76	29.53	100	100
IFCC- BC AU	Photometry UV- RH	7.27	6.89	32.76	32.11	100	100

IFCC- SD	Photometry UV- AA	13.25	7.1	33	32.2	1.7	5
IFCC- SD	Photometry UV- BC AU	13.25	6.63	33	32.82	0	0
IFCC- SD	Photometry UV- RCc	13.25	6.62	33	31	1.4	9.4
IFCC- SD	Photometry UV- RCI	13.25	6.4	33	29.53	0.1	5.2
IFCC- SD	Photometry UV- RH	13.25	6.89	33	32.11	0.1	0.2
Photometry UV- AA	Photometry UV- BC AU	7.1	6.63	32.2	32.82	100	100
Photometry UV- AA	Photometry UV- RCc	7.1	6.62	32.2	31	100	100
Photometry UV- AA	Photometry UV- RCI	7.1	6.4	32.2	29.53	100	100
Photometry UV- AA	Photometry UV- RH	7.1	6.89	32.2	32.11	100	100
Photometry UV- BC AU	Photometry UV- RCc	6.63	6.62	32.82	31	100	100
Photometry UV- BC AU	Photometry UV- RCI	6.63	6.4	32.82	29.53	100	99.8
Photometry UV- BC AU	Photometry UV- RH	6.63	6.89	32.82	32.11	100	100
Photometry UV- RCc	Photometry UV- RCI	6.62	6.4	31	29.53	100	100
Photometry UV- RCc	Photometry UV- RH	6.62	6.89	31	32.11	100	100
Photometry UV- RCI	Photometry UV- RH	6.4	6.89	29.53	32.11	100	100
			ALP				
IFCC- AA	IFCC- BC AU	50.11	50	101.7	113.73	100	9.5
IFCC- AA	IFCC- RCc	50.11	46.15	101.7	83.23	100	5.2
IFCC- AA	IFCC- RCI	50.11	48.47	101.7	88.44	100	11.4
IFCC- AA	IFCC- RCMira	50.11	50.67	101.7	96.83	100	79.6
IFCC- AA	IFCC- RH	50.11	48.48	101.7	96.73	100	68.3
IFCC- AA	IFCC- SD	50.11	48.67	101.7	93.42	100	63.3
IFCC- BC AU	IFCC- RCc	50	46.15	113.73	83.23	100	0
IFCC- BC AU	IFCC- RCI	50	48.47	113.73	88.44	100	0
IFCC- BC AU	IFCC- RCMira	50	50.67	113.73	96.83	99.7	4.5
IFCC- BC AU	IFCC- RH	50	48.48	113.73	96.73	100	0
IFCC- BC AU	IFCC- SD	50	48.67	113.73	93.42	100	0
IFCC- RCc	IFCC- RCI	46.15	48.47	83.23	88.44	100	100
IFCC- RCc	IFCC- RCMira	46.15	50.67	83.23	96.83	97.7	22.1
IFCC- RCc	IFCC- RH	46.15	48.48	83.23	96.73	100	55
IFCC- RCc	IFCC- SD	46.15	48.67	83.23	93.42	100	81.4
IFCC- RCI	IFCC- RCMira	48.47	50.67	88.44	96.83	100	49.4
IFCC- RCI	IFCC- RH	48.47	48.48	88.44	96.73	100	93.8
IFCC- RCI	IFCC- SD	48.47	48.67	88.44	93.42	100	99.9
IFCC- RCMira	IFCC- RH	50.67	48.48	96.83	96.73	99.4	92.9
IFCC- RCMira	IFCC- SD	50.67	48.67	96.83	93.42	99.9	93.8
IFCC- RH	IFCC- SD	48.48	48.67	96.73	93.42	100	98.1

AMY							
IFCC- AA	IFCC- BC AU	52.78	55.58	75.89	77.17	100	100
IFCC- AA	IFCC- RCc	52.78	54.77	75.89	76.54	100	100
IFCC- AA	IFCC- RCI	52.78	56.38	75.89	77.94	100	100
IFCC- AA	IFCC- RH	52.78	52.35	75.89	73.95	100	100
IFCC- AA	CNP-G3- SD	52.78	52.29	75.89	74.17	100	100
IFCC- BC AU	IFCC- RCc	55.58	54.77	77.17	76.54	100	100
IFCC- BC AU	IFCC- RCI	55.58	56.38	77.17	77.94	100	100
IFCC- BC AU	IFCC- RH	55.58	52.35	77.17	73.95	100	100
IFCC- BC AU	CNP-G3- SD	55.58	52.29	77.17	74.17	100	100
IFCC- RCc	IFCC- RCI	54.77	56.38	76.54	77.94	100	100
IFCC- RCc	IFCC- RH	54.77	52.35	76.54	73.95	100	100
IFCC- RCc	CNP-G3- SD	54.77	52.29	76.54	74.17	100	100
IFCC- RCI	IFCC- RH	56.38	52.35	77.94	73.95	100	100
IFCC- RCI	CNP-G3- SD	56.38	52.29	77.94	74.17	100	100
IFCC- RH	CNP-G3- SD	52.35	52.29	73.95	74.17	100	100
AST							
IFCC- BC AU	IFCC- RH	12.43	11	44.7	43.67	100	100
IFCC- BC AU	IFCC- SD	12.43	12.58	44.7	45.33	100	100
IFCC- BC AU	Photometry UV- AA	12.43	10.3	44.7	39.9	100	100
IFCC- BC AU	Photometry UV- BC AU	12.43	11.96	44.7	44.69	100	100
IFCC- BC AU	Photometry UV- RCc	12.43	10.78	44.7	39.56	100	100
IFCC- BC AU	Photometry UV- RCI	12.43	10.6	44.7	39.07	100	100
IFCC- BC AU	Photometry UV- RH	12.43	10.42	44.7	41.42	100	100
IFCC- RH	IFCC- SD	11	12.58	43.67	45.33	100	100
IFCC- RH	Photometry UV- AA	11	10.3	43.67	39.9	100	100
IFCC- RH	Photometry UV- BC AU	11	11.96	43.67	44.69	100	100
IFCC- RH	Photometry UV- RCc	11	10.78	43.67	39.56	100	100
IFCC- RH	Photometry UV- RCI	11	10.6	43.67	39.07	100	100
IFCC- RH	Photometry UV- RH	11	10.42	43.67	41.42	100	100
IFCC- SD	Photometry UV- AA	12.58	10.3	45.33	39.9	100	100
IFCC- SD	Photometry UV- BC AU	12.58	11.96	45.33	44.69	100	100
IFCC- SD	Photometry UV- RCc	12.58	10.78	45.33	39.56	100	100
IFCC- SD	Photometry UV- RCI	12.58	10.6	45.33	39.07	100	100
IFCC- SD	Photometry UV- RH	12.58	10.42	45.33	41.42	100	100
Photometry UV- AA	Photometry UV- BC AU	10.3	11.96	39.9	44.69	100	100
Photometry UV- AA	Photometry UV- RCc	10.3	10.78	39.9	39.56	100	100
Photometry UV- AA	Photometry UV- RCI	10.3	10.6	39.9	39.07	100	100

Photometry UV- AA	Photometry UV- RH	10.3	10.42	39.9	41.42	100	100
Photometry UV- BC AU	Photometry UV- RCc	11.96	10.78	44.69	39.56	100	100
Photometry UV- BC AU	Photometry UV- RCI	11.96	10.6	44.69	39.07	100	100
Photometry UV- BC AU	Photometry UV- RH	11.96	10.42	44.69	41.42	100	100
Photometry UV- RCc	Photometry UV- RCI	10.78	10.6	39.56	39.07	100	100
Photometry UV- RCc	Photometry UV- RH	10.78	10.42	39.56	41.42	100	100
Photometry UV- RCI	Photometry UV- RH	10.6	10.42	39.07	41.42	100	100
CALCIUM							
Arsenaso III- AA	Arsenaso III- BC AU	2.42	2.41	2.39	2.37	93.1	83.6
Arsenaso III- AA	NM-BAPTA- RCI	2.42	2.39	2.39	2.34	96.3	95
Arsenaso III- AA	cresolphthalein - BC AU	2.42	2.38	2.39	2.27	87	47.1
Arsenaso III- AA	cresolphthalein - RCI	2.42	2.42	2.39	2.33	99.2	85
Arsenaso III- AA	o-cresolphthalein - SD	2.42	2.33	2.39	2.25	52.1	63.5
Arsenaso III- BC AU	NM-BAPTA- RCI	2.41	2.39	2.37	2.34	97.6	98.1
Arsenaso III- BC AU	cresolphthalein - BC AU	2.41	2.38	2.37	2.27	79.1	30.3
Arsenaso III- BC AU	cresolphthalein - RCI	2.41	2.42	2.37	2.33	99.9	94.2
Arsenaso III- BC AU	cresolphthalein - SD	2.41	2.33	2.37	2.25	32.8	26
NM-BAPTA- RCI	cresolphthalein - BC AU	2.39	2.38	2.34	2.27	98.9	84.4
NM-BAPTA- RCI	cresolphthalein - RCI	2.39	2.42	2.34	2.33	99.4	99.7
NM-BAPTA- RCI	cresolphthalein - SD	2.39	2.33	2.34	2.25	89.2	94.5
cresolphthalein - BC AU	cresolphthalein - RCI	2.38	2.42	2.27	2.33	92.4	90.6
cresolphthalein - BC AU	cresolphthalein - SD	2.38	2.33	2.27	2.25	91.9	96.7
cresolphthalein - RCI	cresolphthalein - SD	2.42	2.33	2.33	2.25	63.9	93.4
CHLORIDE							
Indirect ISE- AA	Indirect ISE- BC AU	105.27	105.06	102.7	102.71	100	100
Indirect ISE- AA	Indirect ISE- SD	105.27	102.33	102.7	117.67	95.1	0
Indirect ISE- BC AU	Indirect ISE- SD	105.06	102.33	102.71	117.67	80.4	0
CHOLESTEROL							
CHOD-PAP- AA	CHOD-PAP- BC AU	7.54	7.79	6.6	6.73	100	100

CHOD-PAP-AA	CHOD-PAP-RCc	7.54	7.36	6.6	6.5	100	100
CHOD-PAP-AA	CHOD-PAP-RCI	7.54	7.53	6.6	6.61	100	100
CHOD-PAP-AA	CHOD-PAP-RH	7.54	7.38	6.6	6.55	100	100
CHOD-PAP-AA	CHOD-PAP-SD	7.54	7.63	6.6	6.32	100	100
CHOD-PAP-BC AU	CHOD-PAP-RCc	7.79	7.36	6.73	6.5	100	100
CHOD-PAP-BC AU	CHOD-PAP-RCI	7.79	7.53	6.73	6.61	100	100
CHOD-PAP-BC AU	CHOD-PAP-RH	7.79	7.38	6.73	6.55	98.9	98.9
CHOD-PAP-BC AU	CHOD-PAP-SD	7.79	7.63	6.73	6.32	100	95.1
CHOD-PAP-RCc	CHOD-PAP-RCI	7.36	7.53	6.5	6.61	100	100
CHOD-PAP-RCc	CHOD-PAP-RH	7.36	7.38	6.5	6.55	100	100
CHOD-PAP-RCc	CHOD-PAP-SD	7.36	7.63	6.5	6.32	100	100
CHOD-PAP-RCI	CHOD-PAP-RH	7.53	7.38	6.61	6.55	100	100
CHOD-PAP-RCI	CHOD-PAP-SD	7.53	7.63	6.61	6.32	100	100
CHOD-PAP-RH	CHOD-PAP-SD	7.38	7.63	6.55	6.32	100	100
			CK				
IFCC- AA	IFCC- BC AU	52.1	55.09	137.1	136.78	100	100
IFCC- AA	IFCC- RCc	52.1	54.17	137.1	133.33	100	100
IFCC- AA	IFCC - RCI	52.1	55.75	137.1	137.58	100	100
IFCC- AA	IFCC- RH	52.1	53.5	137.1	133.11	100	100
IFCC- AA	IFCC- SD	52.1	52.71	137.1	130.57	100	100
IFCC- BC AU	IFCC- RCc	55.09	54.17	136.78	133.33	100	100
IFCC- BC AU	IFCC - RCI	55.09	55.75	136.78	137.58	100	100
IFCC- BC AU	IFCC- RH	55.09	53.5	136.78	133.11	100	100
IFCC- BC AU	IFCC- SD	55.09	52.71	136.78	130.57	100	100
IFCC- RCc	IFCC - RCI	54.17	55.75	133.33	137.58	100	100
IFCC- RCc	IFCC- RH	54.17	53.5	133.33	133.11	100	100
IFCC- RCc	IFCC- SD	54.17	52.71	133.33	130.57	100	100
IFCC - RCI	IFCC- RH	55.75	53.5	137.58	133.11	100	100
IFCC - RCI	IFCC- SD	55.75	52.71	137.58	130.57	100	100
IFCC- RH	IFCC- SD	53.5	52.71	133.11	130.57	100	100
			CREATININE				
Compensated Jaffe- AA	Compensated Jaffe- BC AU	67.5	60.33	222.83	179.91	29.6	1.9
Compensated Jaffe- AA	Compensated Jaffe- RCc	67.5	64.42	222.83	181.5	96.2	0.1
Compensated Jaffe- AA	Compensated Jaffe- RCI	67.5	66.46	222.83	177.38	100	0
Compensated Jaffe- AA	Compensated Jaffe- RH	67.5	63.36	222.83	197.92	95.4	14.9
Compensated Jaffe- AA	Compensated Jaffe- SD	67.5	63.6	222.83	230	98	89.9

EQA SURVEY 3

Compensated Jaffe- AA	Enzymatic method- BC AU	67.5	60	222.83	151.45	54.9	0
Compensated Jaffe- AA	Non-Compensated Jaffe- BC AU	67.5	72.55	222.83	187.27	91.7	0.2
Compensated Jaffe- AA	Non-compensated Jaffe- RH	67.5	67.12	222.83	199.67	100	21.5
Compensated Jaffe- BC AU	Compensated Jaffe- RCc	60.33	64.42	179.91	181.5	83.4	89.6
Compensated Jaffe- BC AU	Compensated Jaffe- RCI	60.33	66.46	179.91	177.38	49.4	55.2
Compensated Jaffe- BC AU	Compensated Jaffe- RH	60.33	63.36	179.91	197.92	100	12.2
Compensated Jaffe- BC AU	Compensated Jaffe- SD	60.33	63.6	179.91	230	99.9	0
Compensated Jaffe- BC AU	Enzymatic method- BC AU	60.33	60	179.91	151.45	100	0
Compensated Jaffe- BC AU	Non-Compensated Jaffe- BC AU	60.33	72.55	179.91	187.27	0	0
Compensated Jaffe- BC AU	Non-compensated Jaffe- RH	60.33	67.12	179.91	199.67	41.9	42.2
Compensated Jaffe- RCc	Compensated Jaffe- RC	64.42	66.46	181.5	177.38	100	100
Compensated Jaffe- RCc	Compensated Jaffe- RH	64.42	63.36	181.5	197.92	100	78.8
Compensated Jaffe- RCc	Compensated Jaffe- SD	64.42	63.6	181.5	230	100	0
Compensated Jaffe- RCc	Enzymatic method- BC AU	64.42	60	181.5	151.45	99.7	0.8
Compensated Jaffe- RCc	Non-Compensated Jaffe- BC AU	64.42	72.55	181.5	187.27	21	54.5
Compensated Jaffe- RCc	Non-compensated Jaffe- RH	64.42	67.12	181.5	199.67	100	33.1
Compensated Jaffe- RCI	Compensated Jaffe- RH	66.46	63.36	177.38	197.92	100	52.4
Compensated Jaffe- RCI	Compensated Jaffe- SD	66.46	63.6	177.38	230	100	0
Compensated Jaffe- RCI	Enzymatic method- BC AU	66.46	60	177.38	151.45	99.1	29.1
Compensated Jaffe- RCI	Non-Compensated Jaffe- BC AU	66.46	72.55	177.38	187.27	77	89
Compensated Jaffe- RCI	Non-compensated Jaffe- RH	66.46	67.12	177.38	199.67	100	9.1
Compensated Jaffe- RH	Compensated Jaffe- SD	63.36	63.6	197.92	230	100	0
Compensated Jaffe- RH	Enzymatic method- BC AU	63.36	60	197.92	151.45	100	0

Compensated Jaffe- RH	Non-Compensated Jaffe- BC AU	63.36	72.55	197.92	187.27	34.7	72.7
Compensated Jaffe- RH	Non-compensated Jaffe- RH	63.36	67.12	197.92	199.67	100	97.8
Compensated Jaffe- SD	Enzymatic method - BC AU	63.6	60	230	151.45	100	0
Compensated Jaffe- SD	Non-Compensated Jaffe- BC AU	63.6	72.55	230	187.27	37	0
Compensated Jaffe- SD	Non-compensated Jaffe- RH	63.6	67.12	230	199.67	100	6.3
Enzymatic method- BC AU	Non-Compensated Jaffe- BC AU	60	72.55	151.45	187.27	0	2.6
Enzymatic method- BC AU	Non-compensated Jaffe- RH	60	67.12	151.45	199.67	95.5	0
Non-Compensated Jaffe- BC AU	Non-compensated Jaffe- RH	72.55	67.12	187.27	199.67	95.7	65.9
			GGT				
IFCC- AA	IFCC- BC AU	13.64	13.65	60.82	59.64	100	100
IFCC- AA	IFCC- RCc	13.64	13.08	60.82	59.17	100	100
IFCC- AA	IFCC- RCI	13.64	12.33	60.82	60.89	100	100
IFCC- AA	IFCC- RH	13.64	14.54	60.82	58.54	100	100
IFCC- AA	IFCC- SD	13.64	16.42	60.82	60.69	100	100
IFCC- BC AU	IFCC- RCc	13.65	13.08	59.64	59.17	100	100
IFCC- BC AU	IFCC- RCI	13.65	12.33	59.64	60.89	100	100
IFCC- BC AU	IFCC- RH	13.65	14.54	59.64	58.54	100	100
IFCC- BC AU	IFCC- SD	13.65	16.42	59.64	60.69	87.4	87.4
IFCC- RCc	IFCC- RCI	13.08	12.33	59.17	60.89	100	100
IFCC- RCc	IFCC- RH	13.08	14.54	59.17	58.54	100	100
IFCC- RCc	IFCC- SD	13.08	16.42	59.17	60.69	100	100
IFCC- RCI	IFCC- RH	12.33	14.54	60.89	58.54	100	100
IFCC- RCI	IFCC- SD	12.33	16.42	60.89	60.69	90.7	90.7
IFCC- RH	IFCC- SD	14.54	16.42	58.54	60.69	100	100
			GLUCOSE				
GOD-PAP- BC AU	GOD-PAP- RH	3.83	3.76	4.53	4.49	99.5	99.9
GOD-PAP- BC AU	Hexokinase- AA	3.83	3.75	4.53	4.58	100	100
GOD-PAP- BC AU	Hexokinase- BC AU	3.83	3.77	4.53	4.47	100	100
GOD-PAP- BC AU	Hexokinase- RCc	3.83	3.75	4.53	4.47	100	100
GOD-PAP- BC AU	Hexokinase- RCI	3.83	3.76	4.53	4.47	100	100
GOD-PAP- BC AU	Hexokinase- SD	3.83	3.8	4.53	4.81	100	98.4
GOD-PAP- RH	Hexokinase- AA	3.76	3.75	4.49	4.58	100	99.8

GOD-PAP- RH	Hexokinase-BC AU	3.76	3.77	4.49	4.47	98.2	99.4
GOD-PAP- RH	Hexokinase-RCc	3.76	3.75	4.49	4.47	100	100
GOD-PAP- RH	Hexokinase-RCI	3.76	3.76	4.49	4.47	99.8	99.9
GOD-PAP- RH	Hexokinase-SD	3.76	3.8	4.49	4.81	99.8	69.7
Hexokinase-AA	Hexokinase-BC AU	3.75	3.77	4.58	4.47	100	99
Hexokinase-AA	Hexokinase-RCc	3.75	3.75	4.58	4.47	100	100
Hexokinase-AA	Hexokinase-RCI	3.75	3.76	4.58	4.47	100	100
Hexokinase-AA	Hexokinase-SD	3.75	3.8	4.58	4.81	100	99.3
Hexokinase-BC AU	Hexokinase-RCc	3.77	3.75	4.47	4.47	100	100
Hexokinase-BC AU	Hexokinase-RCI	3.77	3.76	4.47	4.47	100	100
Hexokinase-BC AU	Hexokinase-SD	3.77	3.8	4.47	4.81	100	8.9
Hexokinase-RCc	Hexokinase-RCI	3.75	3.76	4.47	4.47	100	100
Hexokinase-RCc	Hexokinase-SD	3.75	3.8	4.47	4.81	100	88.6
Hexokinase-RCI	Hexokinase-SD	3.76	3.8	4.47	4.81	100	65.9
			HDL				
Homogenous-AA	Homogenous-BC AU	1.61	1.61	1.74	1.77	100	100
Homogenous-AA	Homogenous-RCc	1.61	1.64	1.74	1.63	100	100
Homogenous-AA	Homogenous-RCI	1.61	1.71	1.74	1.7	100	100
Homogenous-AA	Homogenous-RH	1.61	1.73	1.74	1.55	100	78.9
Homogenous-AA	Homogenous-SD	1.61	1.7	1.74	1.65	99.3	99.6
Homogenous-BC AU	Homogenous-RCc	1.61	1.64	1.77	1.63	100	84
Homogenous-BC AU	Homogenous-RCI	1.61	1.71	1.77	1.7	100	100
Homogenous-BC AU	Homogenous-RH	1.61	1.73	1.77	1.55	70.4	38.5
Homogenous-BC AU	Homogenous-SD	1.61	1.7	1.77	1.65	66.7	84.3
Homogenous-RCc	Homogenous-RCI	1.64	1.71	1.63	1.7	100	100
Homogenous-RCc	Homogenous-RH	1.64	1.73	1.63	1.55	99.9	100
Homogenous-RCc	Homogenous-SD	1.64	1.7	1.63	1.65	99.7	100
Homogenous-RCI	Homogenous-RH	1.71	1.73	1.7	1.55	100	94.4
Homogenous-RCI	Homogenous-SD	1.71	1.7	1.7	1.65	100	99.6

Homogenous- RH	Homogenous- SD	1.73	1.7	1.55	1.65	99.2	86.2
			IRON				
Ferene- AA	Ferene - HP	19	18.83	39.67	40.5	100	100
Ferene- AA	Ferene- RH	19	19.64	39.67	41.36	100	100
Ferene- AA	Ferene- SD	19	18.82	39.67	40.27	100	100
Ferene- AA	Ferrozine- RCc	19	19.69	39.67	40.77	100	100
Ferene- AA	Ferrozine- RCI	19	20.44	39.67	41.2	100	100
Ferene- AA	Ferrozine- RH	19	20.11	39.67	41.56	100	100
Ferene- AA	TPTZ- BC AU	19	19.11	39.67	41.98	100	100
Ferene- HP	Ferene- RH	18.83	19.64	40.5	41.36	100	100
Ferene- HP	Ferene- SD	18.83	18.82	40.5	40.27	100	100
Ferene- HP	Ferrozine- RCc	18.83	19.69	40.5	40.77	100	100
Ferene- HP	Ferrozine- RCI	18.83	20.44	40.5	41.2	100	100
Ferene- HP	Ferrozine- RH	18.83	20.11	40.5	41.56	100	100
Ferene- HP	TPTZ- BC AU	18.83	19.11	40.5	41.98	100	100
Ferene- RH	Ferene- SD	19.64	18.82	41.36	40.27	100	100
Ferene- RH	Ferrozine- RCc	19.64	19.69	41.36	40.77	100	100
Ferene- RH	Ferrozine- RCI	19.64	20.44	41.36	41.2	100	100
Ferene- RH	Ferrozine- RH	19.64	20.11	41.36	41.56	100	100
Ferene- RH	TPTZ- BC AU	19.64	19.11	41.36	41.98	100	100
Ferene- SD	Ferrozine- RCc	18.82	19.69	40.27	40.77	100	100
Ferene- SD	Ferrozine- RCI	18.82	20.44	40.27	41.2	100	100
Ferene- SD	Ferrozine- RH	18.82	20.11	40.27	41.56	100	100
Ferene- SD	TPTZ- BC AU	18.82	19.11	40.27	41.98	100	100
Ferrozine- RCc	Ferrozine- RCI	19.69	20.44	40.77	41.2	100	100
Ferrozine- RCc	Ferrozine- RH	19.69	20.11	40.77	41.56	100	100
Ferrozine- RCc	TPTZ- BC AU	19.69	19.11	40.77	41.98	100	100
Ferrozine- RCI	Ferrozine - RH	20.44	20.11	41.2	41.56	100	100
Ferrozine- RCI	TPTZ- BC AU	20.44	19.11	41.2	41.98	100	100
Ferrozine- RH	TPTZ- BC AU	20.11	19.11	41.56	41.98	100	100
			LDH				
IFCC- AA	IFCC- BC AU	131.2	133.61	173.56	171.75	100	100
IFCC- AA	IFCC- RCc	131.2	122	173.56	164.86	100	100
IFCC- AA	IFCC- RCI	131.2	135.57	173.56	176.86	100	100
IFCC- AA	IFCC- RH	131.2	126.5	173.56	166.25	100	100
IFCC- BC AU	IFCC- RCc	133.61	122	171.75	164.86	99.1	99.7
IFCC- BC AU	IFCC- RCI	133.61	135.57	171.75	176.86	100	100
IFCC- BC AU	IFCC- RH	133.61	126.5	171.75	166.25	99.9	99.9
IFCC- RCc	IFCC- RCI	122	135.57	164.86	176.86	100	99.9
IFCC- RCc	IFCC- RH	122	126.5	164.86	166.25	100	100
IFCC- RCI	IFCC- RH	135.57	126.5	176.86	166.25	100	100
			PHOSPHATE				
Ammonium- molybdate- AA	Ammonium- molybdate- BC AU	0.8	0.79	1.2	1.17	100	100
Ammonium- molybdate- AA	Ammonium- molybdate- RCc	0.8	0.79	1.2	1.17	100	100
Ammonium- molybdate- AA	Ammonium- molybdate- RCI	0.8	0.81	1.2	1.2	100	100

Ammonium-molybdate- BC AU	Ammonium-molybdate- RCc	0.79	0.79	1.17	1.17	100	100
Ammonium-molybdate- BC AU	Ammonium-molybdate- RCI	0.79	0.81	1.17	1.2	99.8	99.8
Ammonium-molybdate- RCc	Ammonium-molybdate- RCI	0.79	0.81	1.17	1.2	100	100
POTASSIUM							
FES-CC	Indirect ISE-AA	4.15	4.16	3.92	3.96	100	100
FES-CC	Indirect ISE-BC AU	4.15	4.12	3.92	3.91	100	100
FES-CC	Indirect ISE-RCc	4.15	4.21	3.92	4	100	100
FES-CC	Indirect ISE - RCI	4.15	4.18	3.92	3.86	100	100
FES-CC	Indirect ISE-SD	4.15	4.12	3.92	3.85	100	100
Indirect ISE-AA	Indirect ISE-BC AU	4.16	4.12	3.96	3.91	100	100
Indirect ISE-AA	Indirect ISE-RCc	4.16	4.21	3.96	4	100	100
Indirect ISE-AA	Indirect ISE - RCI	4.16	4.18	3.96	3.86	100	100
Indirect ISE-AA	Indirect ISE-SD	4.16	4.12	3.96	3.85	100	100
Indirect ISE-BC AU	Indirect ISE-RCc	4.12	4.21	3.91	4	100	100
Indirect ISE-BC AU	Indirect ISE - RCI	4.12	4.18	3.91	3.86	100	100
Indirect ISE-BC AU	Indirect ISE-SD	4.12	4.12	3.91	3.85	100	100
Indirect ISE-RCc	Indirect ISE - RCI	4.21	4.18	4	3.86	100	100
Indirect ISE-RCc	Indirect ISE-SD	4.21	4.12	4	3.85	100	100
Indirect ISE - RCI	Indirect ISE-SD	4.18	4.12	3.86	3.85	100	100
SODIUM							
FES- CC	Indirect ISE-AA	141.13	140.83	145.55	146.82	100	100
FES- CC	Indirect ISE-BC AU	141.13	140.12	145.55	144.78	100	100
FES- CC	Indirect ISE - RCc	141.13	140.73	145.55	146.09	100	99.9
FES- CC	Indirect ISE - RCI	141.13	141.3	145.55	143.45	100	100
FES- CC	Indirect ISE-SD	141.13	140.7	145.55	143.64	100	100
Indirect ISE-AA	Indirect ISE-BC AU	140.83	140.12	146.82	144.78	100	100
Indirect ISE-AA	Indirect ISE - RCc	140.83	140.73	146.82	146.09	100	100
Indirect ISE-AA	Indirect ISE - RCI	140.83	141.3	146.82	143.45	100	100

Indirect ISE-AA	Indirect ISE-SD	140.83	140.7	146.82	143.64	100	100
Indirect ISE-BC AU	Indirect ISE - RCc	140.12	140.73	144.78	146.09	100	92.1
Indirect ISE-BC AU	Indirect ISE - RCI	140.12	141.3	144.78	143.45	100	100
Indirect ISE-BC AU	Indirect ISE-SD	140.12	140.7	144.78	143.64	100	100
Indirect ISE - RCc	Indirect ISE - RCI	140.73	141.3	146.09	143.45	100	95.9
Indirect ISE - RCc	Indirect ISE-SD	140.73	140.7	146.09	143.64	100	95.7
Indirect ISE - RCI	Indirect ISE-SD	141.3	140.7	143.45	143.64	100	100
BILIRUBIN							
Diazo- AA	Diazo- BC AU	6.6	7.36	19.3	20.47	100	77.9
Diazo- AA	Diazo- HP	6.6	6.5	19.3	19	100	100
Diazo- AA	Diazo- RCc	6.6	5.17	19.3	17.25	100	90.2
Diazo- AA	Diazo- RCI	6.6	5.29	19.3	17.24	100	79.7
Diazo- AA	Diazo- RH	6.6	6.45	19.3	18.68	100	100
Diazo- AA	Diazo- SD	6.6	6.08	19.3	18.15	100	99.7
Diazo- BC AU	Diazo- HP	7.36	6.5	20.47	19	100	100
Diazo- BC AU	Diazo- RCc	7.36	5.17	20.47	17.25	100	5.1
Diazo- BC AU	Diazo- RCI	7.36	5.29	20.47	17.24	92.6	46.3
Diazo- BC AU	Diazo- RH	7.36	6.45	20.47	18.68	100	100
Diazo- BC AU	Diazo- SD	7.36	6.08	20.47	18.15	100	97.7
Diazo- HP	Diazo- RCc	6.5	5.17	19	17.25	100	100
Diazo- HP	Diazo- RCI	6.5	5.29	19	17.24	100	100
Diazo- HP	Diazo- RH	6.5	6.45	19	18.68	100	100
Diazo- HP	Diazo- SD	6.5	6.08	19	18.15	100	100
Diazo- RCc	Diazo- RCI	5.17	5.29	17.25	17.24	100	100
Diazo- RCc	Diazo- RH	5.17	6.45	17.25	18.68	100	100
Diazo- RCc	Diazo- SD	5.17	6.08	17.25	18.15	100	100
Diazo- RCI	Diazo- RH	5.29	6.45	17.24	18.68	100	100
Diazo- RCI	Diazo- SD	5.29	6.08	17.24	18.15	100	100
Diazo- RH	Diazo- SD	6.45	6.08	18.68	18.15	100	100
PROTEINS							
Biuret- AA	Biuret- BC AU	66.2	68.16	59.75	63.04	99.3	66.6
Biuret- AA	Biuret- RCc	66.2	68	59.75	62.38	100	100
Biuret- AA	Biuret- RCI	66.2	67	59.75	61.07	100	100
Biuret- AA	Biuret- RH	66.2	67.61	59.75	63.12	100	97.3
Biuret- AA	Biuret- SD	66.2	69.56	59.75	64.25	100	100
Biuret- BC AU	Biuret- RCc	68.16	68	63.04	62.38	100	100
Biuret- BC AU	Biuret- RCI	68.16	67	63.04	61.07	100	99.5
Biuret- BC AU	Biuret- RH	68.16	67.61	63.04	63.12	100	100
Biuret- BC AU	Biuret- SD	68.16	69.56	63.04	64.25	100	100
Biuret- RCc	Biuret- RCI	68	67	62.38	61.07	100	100
Biuret- RCc	Biuret- RH	68	67.61	62.38	63.12	100	100
Biuret- RCc	Biuret- SD	68	69.56	62.38	64.25	100	100
Biuret- RCI	Biuret- RH	67	67.61	61.07	63.12	100	100
Biuret- RCI	Biuret- SD	67	69.56	61.07	64.25	100	100
Biuret- RH	Biuret- SD	67.61	69.56	63.12	64.25	100	100

EQA SURVEY 3

TRIGLYCERIDES							
GPO-PAP - AA	GPO-PAP- BC AU	0.77	0.77	2.08	2.09	100	100
GPO-PAP - AA	GPO-PAP- BC AU	0.77	0.75	2.08	2.04	100	100
GPO-PAP - AA	GPO-PAP- RCI	0.77	0.77	2.08	2.06	100	100
GPO-PAP - AA	GPO-PAP- RH	0.77	0.77	2.08	2.02	100	100
GPO-PAP - AA	GPO-PAP - SD	0.77	0.69	2.08	2.01	100	100
GPO-PAP- BC AU	GPO-PAP- RCc	0.77	0.75	2.09	2.04	100	100
GPO-PAP- BC AU	GPO-PAP- RCI	0.77	0.77	2.09	2.06	100	100
GPO-PAP- BC AU	GPO-PAP- RH	0.77	0.77	2.09	2.02	100	100
GPO-PAP- BC AU	GPO-PAP - SD	0.77	0.69	2.09	2.01	100	100
GPO-PAP- RCc	GPO-PAP- RCI	0.75	0.77	2.04	2.06	100	100
GPO-PAP- RCc	GPO-PAP- RH	0.75	0.77	2.04	2.02	100	100
GPO-PAP- RCc	GPO-PAP - SD	0.75	0.69	2.04	2.01	100	100
GPO-PAP- RCI	GPO-PAP- RH	0.77	0.77	2.06	2.02	100	100
GPO-PAP- RCI	GPO-PAP - SD	0.77	0.69	2.06	2.01	100	100
GPO-PAP- RH	GPO-PAP - SD	0.77	0.69	2.02	2.01	100	100
URATE							
Uricase- BC AU	Uricase,POD- AA	173.76	176.83	292.77	292	100	100
Uricase- BC AU	Uricase,POD- BC AU	173.76	173.22	292.77	292.28	100	100
Uricase- BC AU	Uricase,POD- RCc	173.76	173.14	292.77	283.62	100	100
Uricase- BC AU	Uricase,POD- RCI	173.76	174	292.77	287.56	100	100
Uricase- BC AU	Uricase,POD- RH	173.76	178.23	292.77	297.77	100	100
Uricase- BC AU	Uricase- RCI	173.76	173.5	292.77	284.86	100	100
Uricase- BC AU	Uricase- RH	173.76	175.38	292.77	293.33	100	100
Uricase- BC AU	Uricase- SD	173.76	157.3	292.77	282.73	94.2	94.2
Uricase,POD- AA	Uricase,POD- BC AU	176.83	173.22	292	292.28	100	100
Uricase,POD- AA	Uricase,POD- RCc	176.83	173.14	292	283.62	100	99.9
Uricase,POD- AA	Uricase,POD- RCI	176.83	174	292	287.56	100	100
Uricase,POD- AA	Uricase,POD- RH	176.83	178.23	292	297.77	100	99.8
Uricase,POD- AA	Uricase- RCI	176.83	173.5	292	284.86	100	100

Uricase,POD-AA	Uricase- RH	176.83	175.38	292	293.33	100	100
Uricase,POD-AA	Uricase- SD	176.83	157.3	292	282.73	59.3	77.2
Uricase,POD-BC AU	Uricase,POD-RCc	173.22	173.14	292.28	283.62	100	100
Uricase,POD-BC AU	Uricase,POD-RCI	173.22	174	292.28	287.56	100	100
Uricase,POD-BC AU	Uricase,POD-RH	173.22	178.23	292.28	297.77	100	100
Uricase,POD-BC AU	Uricase- RCI	173.22	173.5	292.28	284.86	100	100
Uricase,POD-BC AU	Uricase- RH	173.22	175.38	292.28	293.33	99.7	99.7
Uricase,POD-BC AU	Uricase- SD	173.22	157.3	292.28	282.73	77.6	77.6
Uricase,POD-RCc	Uricase,POD-RCI	173.14	174	283.62	287.56	100	100
Uricase,POD-RCc	Uricase,POD-RH	173.14	178.23	283.62	297.77	100	100
Uricase,POD-RCc	Uricase- RCI	173.14	173.5	283.62	284.86	100	100
Uricase,POD-RCc	Uricase- RH	173.14	175.38	283.62	293.33	100	100
Uricase,POD-RCc	Uricase- SD	173.14	157.3	283.62	282.73	100	100
Uricase,POD-RCI	Uricase,POD-RH	174	178.23	287.56	297.77	100	100
Uricase,POD-RCI	Uricase- RCI	174	173.5	287.56	284.86	100	100
Uricase,POD-RCI	Uricase- RH	174	175.38	287.56	293.33	100	100
Uricase,POD-RCI	Uricase- SD	174	157.3	287.56	282.73	100	100
Uricase,POD-RH	Uricase- RCI	178.23	173.5	297.77	284.86	100	100
Uricase,POD-RH	Uricase - RH	178.23	175.38	297.77	293.33	100	100
Uricase,POD-RH	Uricase- SD	178.23	157.3	297.77	282.73	94.8	96.8
Uricase- RCI	Uricase - RH	173.5	175.38	284.86	293.33	100	100
Uricase- RCI	Uricase- SD	173.5	157.3	284.86	282.73	100	100
Uricase- RH	Uricase- SD	175.38	157.3	293.33	282.73	84.6	85.3
			UREA				
Urease,GLDH-AA	Urease,GLDH-BC AU	4.4	4.48	5.63	5.6	100	100
Urease,GLDH-AA	Urease,GLDH-HP	4.4	4.42	5.63	5.55	100	100
Urease,GLDH-AA	Urease,GLDH-RCc	4.4	4.39	5.63	5.55	100	100
Urease,GLDH-AA	Urease,GLDH-RCI	4.4	4.22	5.63	5.39	100	99.8
Urease,GLDH-AA	Urease,GLDH-RH	4.4	4.57	5.63	5.65	99.9	100
Urease,GLDH-AA	Urease,GLDH-SD	4.4	4.6	5.63	5.53	98.7	99.3
Urease,GLDH-BC AU	Urease,GLDH-HP	4.48	4.42	5.6	5.55	100	100

EQA SURVEY 3

Urease,GLDH-BC AU	Urease,GLDH-RCc	4.48	4.39	5.6	5.55	100	100
Urease,GLDH-BC AU	Urease,GLDH-RCI	4.48	4.22	5.6	5.39	72.6	90
Urease,GLDH-BC AU	Urease,GLDH-RH	4.48	4.57	5.6	5.65	100	100
Urease,GLDH-BC AU	Urease,GLDH-SD	4.48	4.6	5.6	5.53	87.7	90.4
Urease,GLDH-HP	Urease,GLDH-RCc	4.42	4.39	5.55	5.55	100	100
Urease,GLDH-HP	Urease,GLDH-RCI	4.42	4.22	5.55	5.39	99.9	100
Urease,GLDH-HP	Urease,GLDH-RH	4.42	4.57	5.55	5.65	100	100
Urease,GLDH-HP	Urease,GLDH-SD	4.42	4.6	5.55	5.53	99.3	99.5
Urease,GLDH-RCc	Urease,GLDH-RCI	4.39	4.22	5.55	5.39	100	100
Urease,GLDH-RCc	Urease,GLDH-RH	4.39	4.57	5.55	5.65	100	100
Urease,GLDH-RCc	Urease,GLDH-SD	4.39	4.6	5.55	5.53	97.3	98
Urease,GLDH-RCI	Urease,GLDH-RH	4.22	4.57	5.39	5.65	82	97.5
Urease,GLDH-RCI	Urease,GLDH-SD	4.22	4.6	5.39	5.53	43.5	55.2
Urease,GLDH-RH	Urease,GLDH-SD	4.57	4.6	5.65	5.53	99.6	99.9

The control sample in the third EQA survey was from another manufacturer than the control sample used in the two previous surveys. The sample corresponds to normal concentration levels respecting the appropriate reference intervals for evaluated analytes, except for creatinine, triglycerides and cholesterol. The concentration ranges of native serum samples (which were used for the comparison) to which the control was compared were also normal, except for cholesterol, where the assessment of similar high concentrations was possible. The commutability of C3/2016 was evaluated for 402 pairwise combinations of MPs.

The control was found to be commutable for all MPs used for measurement of 9 analytes: AMY, AST, cholesterol, CK, iron, LDH, phosphate, potassium and triglyceride. It is also highly commutable (less than 20% MP pairs found to be noncommutable) for GGT, glucose, sodium, proteins, urate and urea.

Commutability of this control is better for HDL. Comparing with previous surveys, where the control was found to be noncommutable with 17/21 and 11/14 MP pairs in EQA survey 1 and EQA survey 2, respectively, C3/2016 was found noncommutable with a substantially lower number of pairwise combination of MPs; 6/15 (40%). On the other hand, the analyte-related

commutability was much worse for calcium, creatinine and AP, with noncommutable MP pairs being as high as 10/15, 34/36, and 18/21 for calcium, creatinine and AP, respectively. The noncommutability of control sample for chloride remains the same as in previous surveys where all MP combinations with SD are noncommutable. SD again is the probable cause of noncommutability issues with ALT, as well as glucose and urate.

When assessing only harmonised MPs for commutability, C3/2016 would also be fully commutable for ALT, chloride, GGT, urea and urate. The number of (non)commutable and (non)harmonised MPs in this survey is presented in contingency Table 19.

Table 19. Contingency table showing the number of commutable/noncommutable and harmonised/nonharmonised MP combinations in the EQA survey 3**.**

	SURVEY 3/2016		
	C	**NC**	**Total**
H	303	57	360
NH	3	39	42
Total	306	96	402

H-harmonised, NH – nonharmonised, C – commutable, NC - noncommutable

4.3 Comparison of commutability results for lyophilised control samples

Table 20 shows the comparison of commutability results of evaluated MP pairs using regression analysis according to the CLSI EP14A2 protocol (83) and the proposed false flagging method for commutability evaluation. The analytes evaluated are some of the most common tests requested in medical biochemical laboratories and are representatives of the various analyte groups: carbohydrates, enzymes, electrolytes, nonprotein nitrogen metabolites and lipids. The evaluation is performed on pairwise combinations of most often used MPs in CROQALM EQA.

Table 20. Comparison of commutability conclusions using false flagging (FF) method and regression analysis (CLSI) for commutability evaluation

MP1	MP2	C1/2016		C2/2016		C3/2016	
		FF	CLSI	FF	CLSI	FF	CLSI
ALT							
Photometry UV - AA	Photometry UV - BC AU	C	C	C	C	C	C
Photometry UV - AA	Photometry UV - RCc	C	C	C	C	C	C
Photometry UV - AA	Photometry UV - RCI	C	C	C	C	C	C
Photometry UV - AA	IFCC, PP - SD	NC	C	excl	C	NC	NC
Photometry UV - BC AU	Photometry UV - RCc	C	C	C	C	C	NC
Photometry UV - BC AU	Photometry UV - RCI	C	NC	C	C	C	C
Photometry UV - BC AU	IFCC, PP - SD	NC	C	excl	C	NC	NC
Photometry UV - RCc	Photometry UV - RCI	C	NC	C	C	C	NC
Photometry UV - RCc	IFCC, PP - SD	NC	C	excl	C	NC	NC
Photometry UV - RCI	IFCC, PP - SD	NC	C	excl	C	NC	NC
AST							
Photometry UV - AA	Photometry UV - BC AU	C	C	C	C	C	C
Photometry UV - AA	Photometry UV - RCc	C	C	C	C	C	C
Photometry UV - AA	Photometry UV - RCI	C	C	C	C	C	C
Photometry UV - AA	IFCC, PP - SD	C	C	C	C	C	C
Photometry UV - BC AU	Photometry UV - RCc	C	C	C	C	C	C
Photometry UV - BC AU	Photometry UV - RCI	C	C	C	C	C	C
Photometry UV - BC AU	IFCC, PP - SD	C	C	C	C	C	C
Photometry UV - RCc	Photometry UV - RCI	C	C	C	NC	C	C
Photometry UV - RCc	IFCC, PP - SD	C	C	C	C	C	C
Photometry UV - RCI	IFCC, PP - SD	C	C	C	C	C	C
CHLORIDE							
Indirect ISE - AA	Indirect ISE - BC AU	C	C	NC	C	C	C
Indirect ISE - AA	Indirect ISE - SD	NC	NC	excl	C	NC	NC
Indirect ISE - BC AU	Indirect ISE - SD	NC	NC	excl	C	NC	NC
CHOLESTEROL							
CHOD-PAP - AA	CHOD-PAP - BC AU	C	C	C	C	C	NC
CHOD-PAP - AA	CHOD-PAP - RCc	C	C	C	NC	C	C
CHOD-PAP - AA	CHOD-PAP - RCI	C	NC	C	C	C	C
CHOD-PAP - AA	CHOD-PAP - SD	NC	NC	NC	NC	C	NC
CHOD-PAP - BC AU	CHOD-PAP - RCc	C	C	C	NC	C	NC
CHOD-PAP - BC AU	CHOD-PAP - RCI	C	C	C	C	C	NC
CHOD-PAP - BC AU	CHOD-PAP - SD	NC	NC	NC	NC	C	C
CHOD-PAP - RCc	CHOD-PAP - RCI	C	C	C	C	C	C
CHOD-PAP - RCc	CHOD-PAP - SD	NC	NC	NC	NC	C	NC
CHOD-PAP - RCI	CHOD-PAP - SD	NC	NC	NC	NC	C	NC
CREATININE							
Compensated Jaffe - AA	Compensated Jaffe - BC AU	C	C	excl	NC	NC	NC
Compensated Jaffe - AA	Compensated Jaffe - RCc	C	C	excl	NC	NC	NC
Compensated Jaffe - AA	Compensated Jaffe - RCI	C	C	excl	NC	NC	NC

MP1	MP2	C1/2016		C2/2016		C3/2016	
		FF	CLSI	FF	CLSI	FF	CLSI
Compensated Jaffe - AA	Compensated Jaffe - SD	C	C	C	NC	NC	C
Compensated Jaffe - BC AU	Compensated Jaffe - RCc	NC	C	excl	NC	NC	C
Compensated Jaffe - BC AU	Compensated Jaffe - RCI	NC	C	excl	NC	NC	C
Compensated Jaffe - BC AU	Compensated Jaffe - SD	C	C	excl	C	NC	NC
Compensated Jaffe - RCc	Compensated Jaffe - RCI	C	C	C	C	C	C
Compensated Jaffe - RCc	Compensated Jaffe - SD	C	C	C	NC	NC	NC
Compensated Jaffe - RCI	Compensated Jaffe - SD	C	C	C	C	NC	NC
GGT							
IFCC - AA	IFCC - BC AU	C	C	C	C	C	C
IFCC - AA	IFCC - RCc	C	C	C	C	C	C
IFCC - AA	IFCC - RCI	C	C	C	C	C	C
IFCC - AA	IFCC - SD	NC	C	C	C	C	C
IFCC - BC AU	IFCC - RCc	C	C	C	C	C	C
IFCC - BC AU	IFCC - RCI	C	C	C	C	C	C
IFCC - BC AU	IFCC - SD	NC	C	NC	C	NC	C
IFCC - RCc	IFCC - RCI	C	C	C	C	C	C
IFCC - RCc	IFCC - SD	NC	C	C	C	C	C
IFCC - RCI	IFCC - SD	NC	C	C	C	NC	C
GLUCOSE							
Hexokinase - AA	Hexokinase - BC AU	C	NC	C	C	C	C
Hexokinase - AA	Hexokinase - RCc	C	C	C	NC	C	NC
Hexokinase - AA	Hexokinase - RCI	C	C	C	C	C	C
Hexokinase - AA	Hexokinase - SD	C	NC	C	NC	C	NC
Hexokinase - BC AU	Hexokinase - RCc	C	NC	C	NC	C	NC
Hexokinase - BC AU	Hexokinase - RCI	C	NC	C	C	C	C
Hexokinase - BC AU	Hexokinase - SD	NC	C	C	NC	NC	C
Hexokinase - RCc	Hexokinase - RCI	C	C	C	C	C	C
Hexokinase - RCc	Hexokinase - SD	C	NC	C	NC	NC	NC
Hexokinase - RCI	Hexokinase - SD	C	NC	C	NC	NC	NC
HDL CHOLESTEROL							
Homogenous-AA	Homogenous-BC AU	NC	NC	NC	NC	C	C
Homogenous-AA	Homogenous-RCc	NC	NC	NC	NC	C	NC
Homogenous-AA	Homogenous-RCI	NC	NC	NC	NC	C	NC
Homogenous-AA	Homogenous-SD	NC	NC	NC	NC	C	NC
Homogenous-BC AU	Homogenous-RCc	NC	C	NC	NC	NC	NC
Homogenous-BC AU	Homogenous-RCI	NC	C	NC	NC	C	NC
Homogenous-BC AU	Homogenous-SD	NC	C	/	NC	NC	NC
Homogenous-RCc	Homogenous-RCI	C	NC	C	NC	C	NC
Homogenous-RCc	Homogenous-SD	C	C	C	C	C	C
Homogenous-RCI	Homogenous-SD	C	C	C	NC	C	NC
POTASSIUM							
Indirect ISE - AA	Indirect ISE - BC AU	C	C	C	C	C	C
Indirect ISE - AA	Indirect ISE - RCc	C	C	C	C	C	C
Indirect ISE - AA	Indirect ISE - RCI	C	/	C	/	C	/

MP1	MP2	C1/2016		C2/2016		C3/2016	
		FF	CLSI	FF	CLSI	FF	CLSI
Indirect ISE - AA	Indirect ISE - SD	C	C	C	C	C	C
Indirect ISE - BC AU	Indirect ISE - RCc	C	C	C	C	C	C
Indirect ISE - BC AU	Indirect ISE - RCI	C	/	C	/	C	/
Indirect ISE - BC AU	Indirect ISE - SD	C	C	C	C	C	C
Indirect ISE - RCc	Indirect ISE - RCI	C	/	C	/	C	/
Indirect ISE - RCc	Indirect ISE - SD	C	C	C	C	C	C
Indirect ISE - RCI	Indirect ISE - SD	C	/	C	/	C	/
SODIUM							
Indirect ISE - AA	Indirect ISE - BC AU	C	C	C	C	C	C
Indirect ISE - AA	Indirect ISE - RCc	C	C	C	C	C	C
Indirect ISE - AA	Indirect ISE - RCI	C	/	C	/	C	/
Indirect ISE - AA	Indirect ISE - SD	C	C	C	C	C	C
Indirect ISE - BC AU	Indirect ISE - RCc	C	C	NC	C	NC	C
Indirect ISE - BC AU	Indirect ISE - RCI	C	/	C	/	C	/
Indirect ISE - BC AU	Indirect ISE - SD	C	C	C	C	C	C
Indirect ISE - RCc	Indirect ISE - RCI	C	/	C	/	C	/
Indirect ISE - RCc	Indirect ISE - SD	C	C	C	C	C	C
Indirect ISE - RCI	Indirect ISE - SD	C	/	C	/	C	/
TRIGLYCERIDES							
GPO-PAP - AA	GPO-PAP - BC AU	C	C	C	/	C	C
GPO-PAP - AA	GPO-PAP - RCc	C	C	C	/	C	C
GPO-PAP - AA	GPO-PAP - RCI	C	C	C	/	C	C
GPO-PAP - AA	GPO-PAP - SD	C	C	C	/	C	C
GPO-PAP - BC AU	GPO-PAP - RCc	C	C	C	/	C	C
GPO-PAP - BC AU	GPO-PAP - RCI	C	NC	C	/	C	NC
GPO-PAP - BC AU	GPO-PAP - SD	C	C	C	/	C	C
GPO-PAP - RCc	GPO-PAP - RCI	C	C	C	/	C	C
GPO-PAP - RCc	GPO-PAP - SD	C	C	C	/	C	C
GPO-PAP - RCI	GPO-PAP - SD	C	NC	C	/	C	NC
UREA							
Urease,GLDH - AA	Urease,GLDH - BC AU	C	C	C	C	C	C
Urease,GLDH - AA	Urease,GLDH - RCc	C	C	C	NC	C	C
Urease,GLDH - AA	Urease,GLDH - RCI	C	C	C	NC	C	C
Urease,GLDH - AA	Urease,GLDH - SD	C	C	C	C	C	C
Urease,GLDH - BC AU	Urease,GLDH - RCc	C	C	C	C	C	C
Urease,GLDH - BC AU	Urease,GLDH - RCI	NC	C	excl	C	NC	C
Urease,GLDH - BC AU	Urease,GLDH - SD	C	C	excl	C	NC	C
Urease,GLDH - RCc	Urease,GLDH - RCI	C	C	C	C	C	C
Urease,GLDH - RCc	Urease,GLDH - SD	C	C	C	C	C	C
Urease,GLDH - RCI	Urease,GLDH - SD	C	C	excl	C	NC	C

C- commutable, NC – noncommutable, excl – excluded form analysis

Both methods for commutability evaluation are showing similar results of full to high commutability of all three EQA control samples for AST, potassium, sodium, triglycerides and urea. High commutability is also confirmed by two methods for ALT in C2/2016, creatinine in C1/2016 and GGT in C2/2016 and C3/2016.

C1/2016 showed moderate commutability to noncommutability for chloride, cholesterol and HDL by both methods. ALT also showed moderate commutability, although the results were different in regression analysis and false flagging method when specific MP pairs are analysed. Both methods also agree on moderate to full noncommutability of C2/2016 for cholesterol and C3/2016 for ALT, chloride and creatinine. The high number of MPs excluded from analysis by false flagging method for chloride and creatinine in EQA survey 2 limits the confirmation of noncommutability of C2/2016 by both methods.

The disagreement on commutability results coming from the regression analysis in CLSI recommended protocol and our method is observed for GGT in C1/2016, HDL and cholesterol in C3/2016 and glucose in all three control samples.

5. DISCUSSION

In the era of harmonisation and standardisation in laboratory medicine, it is very important to recognise and follow all necessary requirements to produce patient result traceable to highest order RMs in order to achieve global comparability and apply universally recommended clinical guidelines. This is why the 'temple of standardisation' is illustratively presented as the temple standing on the 'pillars' of reference measurement system together with the EQA programs offering an assessment of both laboratory and MPs performance, within defined measurement uncertainty. EQA programs that are offering such evaluation, are nowadays gaining more attention since they provide the information on the quality of performance, comparability to the other laboratories and traceability of their results to the reference value.

Commutability is a property of control material related to MPs used to measure a respective analyte. As described in VIM and ISO definitions (37,79), the materials have to show '*the closeness of agreement"* or '*the equivalence of mathematical relationship'* between the results of different MPs as results obtained on patient samples for the same analyte. Because the measurements always include some level of uncertainty, one cannot expect the same results for these two kinds of samples but rather equivalent, or similar results regarding the intended use of control material. Thus, assessment of commutability always includes at least two MPs in the evaluation, and the material is further classified accordingly for these MPs. When one of the MPs is the reference MP and insensitive to commutability issues, the conclusion on commutability can be drawn for the one MP by measurements of both control and patient samples with reference and evaluated MP.

Our first approach in evaluating commutability of EQA control samples was aimed to analysis of the differences of mean MP differences between results of serum and control samples obtained at the same time and under the same analytical conditions. The aim was to investigate different behaviour of MPs on the serum and control samples defined by statistical significance of these differences. The approach was an extension of the work described by Cobbaert et al. (23) where 'spy' serum sample was introduced in the Dutch foundation for Quality Assurance in Clinical Diagnostics (SKML – *Stichting Kwaliteteitsbewaking Medische Laboratoriumdiagnostiek*) for 'sensing' commutability, or as an indication of drifting commutability that has been established. They described this analysis as a pragmatic approach towards commutability assessment, questioning the feasibility of full commutability assessment expected to be scheduled periodically in any EQA scheme. In their approach, the

results from native serum sample and control are compared among MPs, expecting the ratio to be 100% for all evaluated MPs. To investigate whether the difference of 100% is statistically significant, a *t*-test was performed. In a similar way, we evaluated the statistical significance of differences between mean differences among each pairwise combination of MPs using ANOVA analysis (examples in Tables 10, 11 and 12) since many MPs are used for measurement of an analyte, with $P<0.05$ being the limit of significance. We assessed the statistical significance without any connection to clinical limits or APS in the EQA program, and on many occasions, we observed that apparently the same differences between mean differences among results from control and serum samples could lead to opposite commutability decisions (Table 12). When trying to compare these differences to the limits of total error according to biological variation data, we found out that it was hard to decide which proportion of APS should be chosen as commutability criterion. In spite of limitations observed in ANOVA analysis, we feel that it was a valuable approach in the description and evaluation of statistical differences observed for used MPs. Both graphical presentation of data and calculated differences of means between the pairwise comparison of MPs are very informative, giving clear indications for commutability issues (examples in Figures 13, 14 and 15).

Since it has to be proven that the control samples behave as patient samples when measured by two MPs, the regression analysis in the commutability assessment chosen by Eckfeld at all. (94) was the logic approach considering evaluated concentration range of an analyte. The regression analysis is often used in the laboratory when two MPs are compared, and the relationship between measurements on selected MPs defined. This approach can be valuable when assessing the commutability of calibrators, assessing both bias and commutability of materials intended to be used for calibrating field MPs. In an EQA setting, when control materials have to be validated for commutability with many MPs used by participants the comparison of the measurements with reference MP is logistically and economically very demanding. A very strict protocol has to be followed to collect patient serum samples with concentrations of evaluated analyte that spans various concentration levels of control materials. The assessment is further complicated by the fact that the evaluation protocol includes specific reagent lots used at the time of evaluation and the conclusions on commutability might be quite different for various lots, very often changed in the course of laboratory work. In addition to that, the volume of samples required for analysis using reference MP is usually very high, compared to the volumes needed for analysis with routine MP. For example, as opposed to less than 100 µl needed for creatinine analysis by routine MP, the usual volume needed for analysis of creatinine by ID-GC/MS (Isotope Dilution-Gas Chromatography/Mass Spectrometry)

reference MP is 0.2 to 3 ml, depending on the concentration of the analyte. The volumes are even higher for electrolytes, 3×2 ml on average. In addition to high volumes needed for analysis, the logistics of transporting the samples to the reference laboratory might be demanding, especially if there is no such laboratory in the country. The price is another demanding aspect of analysis in a reference laboratory. Starting from the price of about 200 € for analysis of one enzyme in each sample, the price is usually not less than 1000 € for hormones like cortisol (personal communication). Adding all together, the price for at least 20 patient samples and one control sample can be substantially high, especially if many MPs have to be validated for various parameters in the commutability assessment, as usually is the case in an EQA program.

In our attempt to follow the CLSI EP14-A3 (95) recommended evaluation of commutability using Deming regression analysis, we found out that the statistically determined limits for commutability presented as 95% prediction interval around the regression line is very often not an objective criterion for such evaluation. As presented in Figure 11, the width of 95% prediction interval depends on the initial relationship of measurement results but even more on the type of regression analysis chosen for assessment of commutability. The prediction interval around Deming regression line is very restrictive, many times even for the patient results. The percentage of patient results exceeding the limits was 30.1% for HDL, and the number is even much higher for some other analytes or individual pairs of MPs. Thus, the control samples exceeding these limits may belong to the scatter of patient results and because of that show the intraassay characteristics similar to the characteristics of patient samples, yet not in the commutability limits and accordingly, noncommutable. On the other hand, the 95% prediction limits derived from Passing and Bablok regression analysis are too permissive to be set as a commutability criterion. In this respect, we evaluated the commutability according to linear regression analysis, where the percentage of patient results indeed reflected the closest approximation of 95% prediction interval around the regression line, wherein total 2.9% patient results were outside of the limits. The limitation to this approach is that we did not have the error-free MP as a comparative method, although to some extent the error factor is reduced by triplicate measurement of all samples.

Statistical limits are very commonly used in the assessment of commutability of control materials. These limits provide numerical, objective criteria whether the scatter of the control samples around the regression line of two MPs shows statistically significant difference than the scatter of the patient samples with the same MPs. In other words, statistical limits provide

the answer to whether the control samples belong to the same population as clinical samples. Thus, the closeness of agreement or mathematical equivalence is assessed. Yet, the limits are different for each MPs combination, resulting in multiple criteria for commutability assessment. In fact, it is the initial relationship between two MPs that defines the limits, where imprecision and single sample effects can lead to very wide acceptance intervals around the regression line. The limits are not the same for all MPs participating in the EQA assessment and have no relation to actual APS of the provider or clinical needs. Creating such different criteria might result in an unequal judgment of quality achieved by the laboratory, or alignment of MP to the true value of the analyte.

Ricos et al. (102) were the first to evaluate commutability of control material using fixed limits based on the allowed bias from the biological variation data. After the regression analysis, the residuals of control samples were expressed as percentage bias from predicted values as defined by patient samples. Although the evaluation of control materials is still based on the distance of residuals from the regression line, they used one criterion for all MPs under evaluation.

Very recently, the IFCC-WG on commutability (105-107) recognised the problem of unequal, statistically defined commutability criteria. The group suggested the use of unique commutability criterion for all MPs combinations under evaluation. The experimental design was created to assess the difference in bias between patient samples and control samples measured by two MPs under evaluation (106). The '*difference in bias*' approach allows taking the uncertainty of the measurement into consideration at the appropriate concentration range, and both bias and uncertainty of that value are compared to a previously established commutability criterion. The authors recognise that '*closeness of agreement*' of control samples with clinical samples is a relative term, advising the use of commutability limits for control samples as a fraction of total allowed deviation, or APS. Because the quality control samples are analysed in singleton in an EQA survey, both bias and imprecision have to be expected for evaluated MPs, and the acceptance criterion should be chosen accordingly. Such commutability criterion encompasses both the properties of MPs under evaluation and the intended use of RM as the trueness control.

The experimental design proposed by IFCC-WG is still logistically and economically demanding, with even more patient samples and more replicates needed for commutability assessment. Besides carefully choosing patient samples with the concentrations of analyte close to the concentrations of evaluated controls, the MPs evaluated must have satisfactory precision profiles. In fact, the protocol is rather restrictive to MPs with adequate precisions in order not

to jeopardise commutability conclusion due to large uncertainty intervals and thus inconclusive conclusions. It can be seen in Figure 7 that depending on the MPs combinations assessed (MP y vs. MP x and MP z vs. MP x), and the different precision profiles regarding individual MPs in the presented example, the conclusions on commutability of RMs are different. RM1, RM2 and RM5 showing commutability in the comparison of MPs y and x, are characterised as 'inconclusive' with the MPs z and x due to the large uncertainty intervals of calculated difference in bias between those MPs. Defining satisfactory precision requires statistical power analysis and depends on the observed closeness of agreement between patient and control samples in the experiment. All these requirements make the commutability assessment of control samples used in an EQA program very hard to perform for even limited number of MPs. For example, commutability assessment for one analyte and five MPs used in the laboratories requires at least 30 patient samples to be measured in triplicate, at the same time on all MPs. In total, at least 450 measurements should be performed on patient samples, or even more if the results show the unequal precision profiles leading to large uncertainty intervals and inconclusive commutability decisions. The required volumes, logistics and prices for such experiment are obviously very hard to follow. The recommended IFCC protocol may give the final answer on commutability of control materials, but it still waits for the confirmation in practice. Besides complexity, the protocol lacks clear guidance for EQA providers on the needed '*fraction of total APS*' to be used as a commutability criterion. It is to be expected that in the absence of clear definitions, various EQA programs will choose different fractions of allowed deviations as the commutability limits potentially resulting in different conclusions on commutability of evaluated control materials for the same MPs. Nevertheless, the IFCC-WG protocol made a large contribution in the reasoning that the '*closeness of agreement*' and '*mathematical equivalence*' have to be observed related to the intended use of processed materials. If the processed material is a calibrator, then the intended use for the calibrator is justifying MPs to meet metrological hierarchy leading to the traceability of results to the higher order RMs. If the calibrator succeeds to harmonise the results from two MPs within defined limits, then it can be considered commutable, and fit for the intended use. As for control samples, the intended use is met if the controls can be used for assessment of laboratory and MP performance. Using such control material, all laboratories and MPs under evaluation should have substantially equal chance to meet the predefined limits. In addition, the performance of individual laboratory and MP on control samples should be equal to the performance seen on patient samples.

Considering requirements, the evaluation of commutability of control materials should be performed in a way that appreciates APS recommended for an analyte. APS should, in turn, reflect the clinical needs or quality standards needed to assure reliable and traceable results for clinical samples. When APS for an analyte are wide with respect to the state-of-the-art variability, then the commutability criterion is also going to be more permissive than the commutability criterion for analytes with very strict APS. If closeness of agreement is assessed, then one has to require that control samples provide the same answer regarding flagging of a laboratory in an EQA survey as for human samples. This is why our approach was extended to the evaluation of difference in flagging of laboratories within the EQA scheme as the basis for the commutability evaluation of control materials used in the same scheme. In this way, we were able to clearly analyse the '*closeness of agreement*' between serum and control samples related to the APS of a scheme. Assuming clinical relevance of the established APS, the commutability limits are then defined by clinical relevance of the information gained from two kinds of samples.

In every EQA survey it can be suspected that certain laboratories are going to be flagged, not being able to satisfy the predefined APS. If the same proportion of laboratories is flagged when serum sample is used as control sample compared to the flagging rate with lyophilised control samples, then one has to assume the same behaviour of those samples in its intended use. If serum and lyophilised control samples are analysed in the same run on the instrument, then one would expect that approximately the same proportion of laboratories be flagged on serum and control samples. There is no reason for a number of laboratories to pass the predefined limits on serum samples, and not pass those limits when lyophilised control samples are used in an EQA survey if the concentrations assessed are approximately at the same level. If this is observed, one must question the properties of lyophilised control samples to act as a substitute for appropriate clinical samples. In other words, one must suspect the commutability of the processed material. As presented in our analysis from results of the EQA survey 1/2016 (Table 14), the majority of laboratories using CHOD-PAP method on Siemens Dimension instrument for cholesterol measurement passed the predefined limits when serum sample was used for comparison between MPs. On the contrary, most laboratories had not passed this limit when lyophilised control sample was used for comparison. What would be the cause of this difference? As it can be seen from Table 14, the results from CHOD-PAP-SD are comparable, or harmonised, to all other evaluated MPs when measurements are performed on serum samples. Yet, the results from CHOD-PAP-SD are substantially different from other MPs when measurements are performed on lyophilised controls, causing a relevant number of laboratories

to be flagged. Thus, one must conclude that the root cause for this observation is not the supposedly poor quality of results from these laboratories, but the potential noncommutability of control material, not giving them the equal chance to meet the predefined APS. The control samples have to be able to demonstrate the agreement between laboratories and to point out '*the bad performers*', but it has to be proven that the control sample is really fit for this purpose, reflecting the behaviour of both laboratories and MPs on clinical samples. Thus, the flagging rate observed in serum should be the same as the flagging rate of laboratories on lyophilised control samples when assessing approximately the same concentration levels. Even if two MPs show different results on serum samples (nonharmonised MPs), one should observe the same pattern of flagging on lyophilised control samples. The difference observed in serum samples should also be present on lyophilised control samples.

The logic of the difference between the false flagging rate for control and fresh serum samples was the basis of our approach in evaluating the commutability of lyophilised control samples. The closeness of agreement between serum and processed material is evaluated by means of assessment of laboratory and MPs performance within EQA survey. The method was named *false flagging method* since it classifies commutability of processed materials according to the observed difference in the flagging rate of laboratories with those materials and serum samples. This logic is similar to the recommendations from IFCC-WG on commutability, stating that the criterion for commutability should be '*a fraction of bias component of acceptance limits for evaluating an EQA or trueness control result*' (105). Thus, the limits for commutability are fixed and connected to the APS in the EQA scheme, allowing the results of EQA control samples to be different from clinical samples only a fraction of total allowance.

Applied into the false flagging method, we set the allowed difference in changing flagging rate to 20% points. The limit is arbitrary and prone to changes depending on the scheme design, number of participants and clinical relevance of the analyte. Our decision was based on the fact, that when evaluating the performance of two sets of results in the corresponding frequency curves, the change in the flagging of results can be expected beyond the APS evaluation limit.

Because the MP groups in CROQALM are relatively small, with many groups not exceeding 10 participants when both analytical method and instrument are chosen as the basis for group differentiation, the probabilities of change in flagging rate are expected to be variable. Hence, the bootstrapping technique is used in order to calculate the probability that the limits are ever going to be exceeded. Just like the sample is drawn from the whole population, the bootstrapped sample is drawn from the original sample or set of results. Starting from the original sample of size N, bootstrapped samples (usually 1000) of the equal size can be generated and the statistics

performed on each sample pooled together, constructing a sampling distribution which can be used to make statistical inference.

If two MPs are harmonised, the rate of flagged results when the groups are joined and evaluated according to the unique target value, as opposed to the flagging rate of laboratories when the groups are split, is within the predefined limit of 20% points. In Tables 14, 16 and 18, harmonised MP pairs are presented with 95-100% harmonisation, or 95-100% of samples not exceeding the predefined limit. Indeed, defining harmonisation with the same logic of flagged results before and after the groups are split can assure that the evaluation of individual laboratories will not be significantly different. Assuming their harmonisation, one can easily observe the different behaviour of those same MPs on the EQA control samples. As seen in the case of CHOD-PAP-SD MP in the EQA survey 1 and 2 (Tables 14 and 16), the flagging rate changed substantially for all combinations of MPs with CHOD-PAP SD, causing at some instances none of the samples to meet the predefined limits. Percentages of commutability for these MPs combinations range from 0.2 – 80.4% and 0.0 – 21.6% for C1/2016 and C2/2016 control samples, respectively. The conclusion on noncommutability of these EQA samples is rather straightforward, after observing quite different behaviour of evaluated MPs with these controls as opposed to their behaviour on serum samples. The rate of falsely flagged results of participating laboratories indicated the commutability issues of these controls. On the other hand, the control C3/2016 showed commutability with all MPs for cholesterol, where the rate of flagged results on serum samples and this control was approximately the same, or within the predefined limits. The conclusions on commutability using false flagging method for cholesterol were the same as conclusions from simple linear regression analysis for C1/2016 and C2/2016, but quite opposite for C3/2016 (Table 20). A statistically significant difference in the behaviour of C3/2016 was found using CLSI protocol for evaluating commutability, although this significance was not observed in terms of the difference of flagging status of laboratories when evaluated according to the same target value. Respecting these findings, we concluded that the behaviour of this control regarding predefined APS for cholesterol was not significantly different in control and serum sample, which makes C3/2016 reliable control for assessing both laboratory and MP performance. The observed difference compared to patient samples in the CLSI protocol might be statistically significant, mainly influenced by the initial relationship between evaluated MPs, yet not significantly different to jeopardize intended use of RM.

The noncommutability of control materials for HDL was confirmed using the false flagging method. In all three surveys, harmonised MPs yield quite different results on lyophilised

controls, although the C3/2016 shows better commutability compared to the other two controls. According to our classification of commutability, C1/2016 and C2/2016 would be described as noncommutable materials for HDL cholesterol with more than 60% noncommutable MP pairs, whereas the control C3/2016 shows moderate noncommutability with 40% noncommutable MP pairs. The results partially agree with conclusions from protocol according to CLSI on commutability for evaluated MPs (Table 20), but some differences are important. As it can be seen in Table 14, C1/2016 shows commutability with almost all MPs including Homogenous-BC AU because the 95% prediction interval around the regression line for all MP pairs is very wide and permissive in terms of commutability decisions (see Figure 12B). Although the scatter of results from this control indeed belongs to the limits of the scatter observed on patient samples, these limits are very wide according to limits allowed in the EQA program. Because of these differences, many laboratories would be falsely flagged if methods are joined and evaluated according to the consensus target value. In conclusion, the statistical equality observed in regression analysis is not restrictive enough for applied APS or clinical needs for this analyte. On the other hand, observed statistical differences in C3/2016 are too insignificant according to APS at this concentration level (the highest concentration of HDL in control samples; 1.8 mmol/L) to jeopardize evaluation in the EQA program, hence the conclusion on commutability differs from the CLSI protocol. In addition, the evaluated MPs in CLSI protocol did not include Roche Hitachi instrument, where pairwise combinations of MPs with this instrument showed noncommutability according to the false flagging method.

Glucose is another example where the APS in the EQA scheme allow the results to differ in a way that the statistically significant differences observed in regression analysis don't influence the decision on the flagging of individual laboratories. The commutability evaluation of control materials according to CLSI protocol results in a high number of noncommutability decisions, while using the false flagging method these materials are considered commutable for the majority of MP combinations (Table 20).

The approach in evaluation commutability using regression analysis used in CLSI recommended protocol is very different than the approach used in the false flagging method. While one looks at statistically significant differences between results from patient and control samples using regression analysis in describing *closeness of agreement*, the false flagging method describes this closeness according to the intended use of control material using fixed limits for commutability which are dependent on the APS used in the EQA program. Respecting these different approaches, we feel that the comparison with commutability results

from CLSI protocol might only serve for explanatory purposes in finding the reason for observed differences, as shown in the cases of HDL and glucose.

As stated before, the commutability decision using false flagging method for harmonised MPs is relatively easy because one starts from the results that are comparable, or substantially equal. For nonharmonised MPs, the decisions on noncommutability are somewhat different, because the results are different even on serum samples. Therefore, in order to prove noncommutability, it has to be shown that the results are even more different on control samples or, that they are comparable, which is also the proof of noncommutability, because the control sample shows different behaviour from a serum sample. Actually, in order to prove commutability of control materials, one would have to prove how '*equally unequal*' are the results on serum and control samples. In our opinion, the logic of analysing the change in flagging rate qualifies for this purpose, because MPs are not only evaluated by the proportion of flagged laboratories, but rather the change in this proportion. Yet, it has to be noticed, that among nonharmonised MPs, many pairs of MPs show noncommutability as well. Nonharmonisation was most likely the cause for noncommutability of some MP pairs, for example, all combinations with IFCC-SD MP for ALT in the EQA survey 1 (Table 14) and almost all nonharmonised MP pairs used to measure calcium in EQA survey 3 (Table 18). Actually, all nonharmonised MPs with less than 80% harmonisation show noncommutability. Evaluation of commutability of nonharmonised MPs is thus considered as a possible limitation of false flagging method in evaluating commutability. The number of noncommutability decisions from nonharmonised MP pairs is shown in Tables 15, 17 and 19, following analysis of each EQA survey.

In fact, it was probably the issue of nonharmonisation resulting in noncommutability the root-cause of exclusion of many MP pairs from the evaluation in the EQA survey 2. In this survey, in order to achieve high concentration levels of some analytes, the native fresh serum was spiked with glucose, urea, bilirubin, sodium and chloride. Since one cannot assume that any processed material behaves like fresh serum sample, we evaluated this kind of material in the third EQA survey, where one volume of the serum was distributed native and the other spiked with these analytes. Both sera were measured in the same run on the instruments, and the false flagging method was used to evaluate the commutability of the spiked volume of serum in order to prove that spiked serum can be used as a substitute for fresh serum samples. After analysis, the total of 48/379 MPs were excluded from further evaluation because of the elevated number of flagged results in spiked serum samples (Appendix). Although the cause for exclusion of many MPs might have been nonharmonisation of MPs, the reason why so many

MPs were excluded for creatinine was unclear to us. The creatinine was not the analyte used for spiking, and the number of excluded harmonised MPs was 16/28. Thus, the assumption that spiking with low amounts of relatively simple analytes assures commutability might not be true, and it must always be checked. We proved that this assumption certainly was not true in the cases of many MPs used for measurement of chloride and bilirubin. The choice of spiking serum in order to achieve high concentration was a good decision for glucose, urea and sodium, but regarding bilirubin and chloride, the pooling of the samples with high concentrations of evaluated analytes might have been the better choice. Nevertheless, the pools of serum samples also have to be validated for commutability with native clinical samples.

The variability of results in the MPs' groups also influences the commutability decisions, regardless of the MPs used for evaluation. Recognising this observation, IFCC-WG requires only MPs with satisfactory precision to participate in the commutability evaluation. High imprecision of the MP expressed as the variability of measurement results around the mean value introduces larger uncertainty estimates which, in turn, can result in too permissive prediction intervals around the regression line in CLSI protocol. In the IFCC-WG recommended protocol, it is expected to result in large uncertainty intervals around the difference in bias between two MPs, making more commutability decisions as inconclusive. As far as our false flagging method is concerned, the variability of results reported for each MP influences the commutability in a way that some materials appear to be noncommutable with MPs having larger variability. The variability that is presented in EQA results might be thought as maximum imprecision for an MP, whose positive influence on nonharmonisation of MPs might be followed by noncommutability decision. When only mean values and difference of means for serum and control samples are observed in the pairwise combination of MPs, it might be seen on some occasions that for same differences we get quite a different harmonisation and commutability decisions, which is, in fact, due to different variabilities between methods. For example, this can be seen in the case of comparison between FES-CC-Indirect-AA and FES-CC-Indirect-SD for potassium measurement in the EQA survey 1 (Table 14). The means and the differences between these two MPs combinations are approximately the same, but the conclusions on commutability are quite different (88% commutability as opposed to 99.9% commutability). The reason is that the standard deviation for Indirect ISE-AA is quite larger (SD = 0.12) than for Indirect ISE-SD (SD = 0.06), and thus more laboratories are flagged in the MP group of Indirect-AA. Similar observations can be seen throughout all three surveys. Regardless of the means, variability means that the substantial number of laboratories have results on the extremities of their distribution, being characterised as ones exceeding the

predefined limits. Moreover, if the limits are very strict, as it is the case with calcium, one can expect that larger variability in the groups would result in more results being flagged, further described as nonharmonisation or noncommutability.

Commutability is an important characteristic of EQA samples enabling evaluation of laboratory and MP performance according to unique target values. Control samples that mimic clinical patient samples can give valuable information on quality standards met in the MBLs as well as harmonisation and standardisation of MPs used in the clinical setting. EQA programs using commutable control samples and reference target values offer multiple evaluation capabilities and are recognised as the most useful programs considering the information they provide to their participants and healthcare community. In order to fulfill the requirements for commutability assessment in the EQA scheme with many MPs used for analysis, we presented a method that can give the EQA providers information on the commutability of MPs used in the program. The method is based on the 'closeness of agreement' of performance of both laboratories and MPs on serum and control samples. This way the intended use of RM in identifying poor performance and/or harmonisation and standardisation of MPs used by the laboratories is recognised. For a commutable control sample, the equal performance of laboratories is expected on patient serum samples and control samples. If the proportion of flagged laboratories changed substantially in only one sample, noncommutability of control materials is assumed, where the behaviour of processed material is different from the native serum sample. We defined the limit of 20% points change in flagging status as the limit for commutability, respecting variability of results within groups and a small number of results for some MPs. By doing so, we were able to demonstrate noncommutability of all lyophilised control materials for some analytes. The number of MPs showing noncommutability with evaluated control samples varied depending on both sample and analyte. The use of the false flagging method can even give substantial evidence in identifying controls where some MPs show a very low percentage of commutability, thus very different behaviour of those samples than patient samples. As shown in Tables 14, 16 and 18, very low commutability percentages for many MPs used for measuring HDL in all three control samples, or AP and creatinine in C3/2016, give a clear conclusion on commutability of those samples. On the other hand, some commutability percentages are in the line of 90%, giving a possibility that near-commutability may be expected, which may require additional analysis using, for example, the IFCC-WG protocol for evaluation. Actually, this protocol may also be advised in cases where only a few MPs show noncommutability, especially if the percentages of commutability are rather high,

raising the need for additional analysis. On the contrary, when controls are defined as showing high noncommutability (noncommutable for > 60 % MPs combinations used in the measurement of stated analyte) or full commutability, there is no need for additional analysis and further costs.

In the proposed commutability evaluation experiment, we were able to analyse both normal and high concentration levels of some analytes in both samples (glucose, urea, cholesterol, sodium, potassium, chlorides and bilirubin). However, the spiking of serum samples with simple analytes similar to the protocol described by SKML scheme where commutability was demonstrated was not so successful in our study. For example, spiking with chloride introduced commutability issues with serum sample, as well as spiking with ditauro-bilirubin for many MPs. In addition, spiking seems to be a problem also for some rather non-specific MPs used for measurement of creatinine, although creatinine was not added to the serum samples. Because one cannot *a priori* assume commutability of any processed material, we would advise using pools of clinical samples for commutability assessment. Pools with a high concentration of several analytes can be collected relatively easy, for example, high levels of urea, creatinine and potassium, since high concentration of these analytes is present in the same clinical entities (chronic kidney failure). In a similar way pools with high concentrations of glucose and lipids can be obtained, high activities of some enzymes, etc. The volumes needed for analysis are also small, and several different pools can be sent together with control samples to participating laboratories in order to evaluate their commutability with MPs in routine use. Once established, the false flagging method can be used periodically in order to evaluate the commutability of existing control materials with new MPs used by participants or to validate the conclusions on commutability across any time period. The proposed commutability criterion can also be adjusted to any specific circumstance of EQA program or clinical significance of the analyte. The price and logistics applied are far more affordable and acceptable, with the criteria of commutability being clearly connected to defined quality requirements in laboratory medicine.

6. CONCLUSIONS

Result of this research showed the following:

- Statistically determined commutability limits using regression analysis offer a numeric and objective assessment of commutability of control materials but are very dependent on the variability of MPs. The commutability limits derived from the 95% prediction interval in regression analysis are highly influenced by precision profiles of evaluated MPs.
- The regression analysis showed that all three control samples are highly noncommutable for evaluated pairwise combinations of MPs used for measurement of cholesterol, HDL cholesterol and glucose. The controls were also found highly noncommutable for chloride at normal concentration level (C1/2016 and C3/2016) and creatinine at high concentration level (C2/2016 and C3/2016). C3/2016 showed to be noncommutable for 60% MP pairs used for measuring ALT.
- Commutability of control materials for all evaluated MP pairs was proven for potassium, sodium, GGT, triglycerides and AST using regression analysis.
- Commutability of control materials might be assessed through an EQA program by analysis of native serum samples and lyophilised control materials at the same time using appropriate MPs.
- Assessment of the statistical significance of the difference of mean MP differences of control and serum samples using ANOVA analysis for commutability evaluation is highly dependent on the number and variability of the data in each MP group and cannot be suggested for commutability assessment within EQA program.
- Commutability evaluation of control materials using statistically determined commutability limits have no association to APS used in the EQA program or clinical relevance of an analyte.
- The *closeness of agreement* between patient samples and control materials can be assessed through evaluation of the flagging rate of laboratories on serum samples and control materials.
- Criteria for commutability limits using false flagging method can be related to APS of the EQA scheme, expecting the similar proportion of laboratories to be flagged in both serum samples and lyophilised control materials under the presumption of comparable concentration levels

- Criteria for commutability limits can be adjusted to specific characteristics of the EQA program (number and variability of data) but also to the clinical significance of analyte.
- Pairwise combinations of MPs involving instrument Siemens Dimension were often found noncommutable using false flagging method.
- Nonharmonised MP combinations very often show noncommutability for control samples; the noncommutability of nonharmonised MP pairs was > 80% in all three controls.
- Commutability of spiked serum samples cannot be presumed: all processed materials have to be evaluated for commutability for all assessed analytes (spiked and not-spiked) and pairwise combination of MPs.
- The false flagging method represents a new approach in commutability evaluation and it can be used for evaluating commutability of control samples within the EQA program of medical biochemical laboratories.
- Using the false flagging method in evaluating commutability, the commutability limits are equally designed for all MP combinations. The limits are connected to established APS of the EQA program and may be set to reflect the clinical relevance of the analyte.

7. REFERENCES

1. Müller MM. Quality and diagnostic perspectives in laboratory diagnostics. Biochem Med. 2010;20(2):144–6.
2. Plebani M. The future of clinical laboratories: more testing or knowledge services? Clin Chem Lab Med. 2005;43(9):893–6.
3. Hawkins R. Managing the Pre- and Post-analytical Phases of the Total Testing Process. Ann Lab Med. 2012;32(1):5–16.
4. CLIA Law & Regulations. Available from: https://wwwn.cdc.gov/clia/Regulatory/default.aspx. Accessed Aug 3rd, 2017.
5. Hrvatska komora medicinskih biokemičara – Hrvatska komora medicinskih biokemičara. Available from: http://www.hkmb.hr/. Accessed Aug 3rd, 2017.
6. International Organisation for Standardisation 15189:2012. Medical laboratories: particular requirements for quality and competence. Geneva, Switzerland: International Organisation for Standardisation, 2007.
7. Revision of the "Guideline of the German Medical Association on Quality Assurance in Medical Laboratory Examinations – Rili-BAEK". LaboratoriumsMedizin. 2015;39(1):26–69.
8. Plebani M. External quality assessment programs: Past, present and future. Jugosl Med Biohemija. 2005;24(3):201–206.
9. Uldall A. Quality assurance in clinical chemistry. Scand J Clin Lab Investig Suppl. 1987;187:1–95.
10. Belk WP, Sunderman FW. A survey of the accuracy of chemical analyses in clinical laboratories. Am J Clin Pathol. 1947;17(11):853–861.
11. Organization WH, others. WHO manual for organizing a national external quality assessment programme for health laboratories and other testing sites. 2016 Available from: http://apps.who.int/iris/bitstream/10665/250117/1/9789241549677-eng.pdf. Accessed Aug 4th, 2017.
12. Sciacovelli L, Secchiero S, Zardo L, Plebani M. The role of the External Quality Assessment. Biochem Med. 2010;20(2):160–4.
13. Kristensen GBB, Meijer P. Interpretation of EQA results and EQA-based trouble shooting. Biochem Med. 2017;27(1):49–62.
14. Jones GRD. The role of EQA in harmonization in laboratory medicine – a global effort. Biochem Med. 2017;27:23–9.
15. Jones GR, Koetsier SD. RCPAQAP First Combined Measurement and Reference Interval Survey. Clin Biochem Rev. 2014;35(4):243–50.
16. Nikolac N, Lenicek Krleza J, Simundic A-M. Preanalytical external quality assessment of the Croatian Society of Medical Biochemistry and Laboratory Medicine and CROQALM: finding undetected weak spots. Biochem Med. 2017;27(1):131–43.
17. International Organisation for Standardisation/International Electrotechnical Comission 17043. Conformity assessment: general requirements for proficiency testing. Geneva, Switzerland: International Organisation for Standardisation, 2010.
18. Lippi G, Plebani M, Simundic A-M. Quality in laboratory diagnostics: from theory to practice. Biochem Med. 2010;20(2):126–30.
19. Bukve T, Stavelin A, Sandberg S. Effect of Participating in a Quality Improvement System over Time for Point-of-Care C-Reactive Protein, Glucose, and Hemoglobin Testing. Clin Chem. 2016;62(11):1474–81.

20. Bhat V, Chavan P, Naresh C, Poladia P. The External Quality Assessment Scheme (EQAS): Experiences of a medium sized accredited laboratory. Clin Chim Acta. 2015;446:61–3.
21. Greaves RF. The central role of external quality assurance in harmonisation and standardisation for laboratory medicine. Clin Chem Lab Med. 2017;55(4):471–3.
22. Ceriotti F. The role of External Quality Assessment Schemes in Monitoring and Improving the Standardization Process. Clin Chim Acta. 2014;432:77–81.
23. Cobbaert C, Weykamp C, Franck P, de Jonge R, Kuypers A, Steigstra H, et al. Systematic monitoring of standardization and harmonization status with commutable EQA-samples—five years experience from The Netherlands. Clin Chim Acta. 2012;414:234–240.
24. Greenberg N. Update on current concepts and meanings in laboratory medicine - Standardization, traceability and harmonization. Clin Chim Acta. 2014;432:49–54.
25. Armbruster D. Metrological Traceability of Assays and Comparability of Patient Test Results. Clin Lab Med. 2017;37(1):119–35.
26. KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Available from: http://www.kdigo.org/clinical_practice_guidelines/pdf/CKD/KDIGO_2012_CKD_GL.pdf. Accessed Aug 14th, 2017.
27. Plebani M. Harmonization in laboratory medicine: the complete picture. Clin Chem Lab Med. 2013;51(4):741–51.
28. Beastall GH. Adding value to laboratory medicine: a professional responsibility. Clin Chem Lab Med. 2013;51(1):221–7.
29. Miller WG, Tate JR, Barth JH, Jones GRD. Harmonization: the Sample, the Measurement, and the Report. Ann Lab Med. 2014;34(3):187–97.
30. Aarsand AK, Sandberg S. How to achieve harmonisation of laboratory testing -The complete picture. Clin Chim Acta Int J Clin Chem. 2014;432:8–14.
31. Panteghini M. Traceability as a unique tool to improve standardization in laboratory medicine. Clin Biochem. 2009;42(4–5):236–40.
32. Koumantakis G. Traceability of measurement results. Clin Biochem Rev. 2008;29(Suppl 1):S61–6.
33. Directive 98/79/EC of the European Parliament and of the Council of 27 October 1998 on in vitro diagnostic medical devices. 1998;L331:1–37.
34. International Organisation for Standardisation Guide 35. Reference materials: general and statistical principles for certification. Geneva, Switzerland, 2006.
35. Thienpont LM, Van Uytfanghe K, De Grande LAC, Reynders D, Das B, Faix JD, et al. Harmonization of Serum Thyroid-Stimulating Hormone Measurements Paves the Way for the Adoption of a More Uniform Reference Interval. Clin Chem. 2017;63(7):1248–60.
36. Warnick GR, Kimberly MM, Waymack PP, Leary ET, Myers GL. Standardization of Measurements for Cholesterol, Triglycerides, and Major Lipoproteins. Lab Med. 2008;39(8):481–90.
37. International Organisation for Standardisation 17511. Metrological traceability of values assigned to calibrators and control materials. In vitro diagnostic medical devices – measurement of quantities in biological samples. Geneva, Switzerland, 2003.
38. Beastall GH, Brouwer N, Quiroga S, Myers GL, prepared on behalf of the Joint Committee for Traceability in Laboratory Medicine. Traceability in laboratory medicine: a global driver for accurate results for patient care. Clin Chem Lab Med. 2017;55(8):1100–8.

39. Thienpont LM, Van Uytfanghe K, Rodríguez Cabaleiro D. Metrological traceability of calibration in the estimation and use of common medical decision-making criteria. Clin Chem Lab Med. 2004;42(7):842–50.
40. Braga F, Panteghini M. Verification of in vitro medical diagnostics (IVD) metrological traceability: Responsibilities and strategies. Clin Chim Acta. 2014;432:55–61.
41. Braga F, Infusino I, Panteghini M. Performance criteria for combined uncertainty budget in the implementation of metrological traceability. Clin Chem Lab Med. 2015;53(6):905–12.
42. Stepman HCM, Tiikkainen U, Stockl D, Vesper HW, Edwards SH, Laitinen H, et al. Measurements for 8 Common Analytes in Native Sera Identify Inadequate Standardization among 6 Routine Laboratory Assays. Clin Chem. 2014;60(6):855–63.
43. White GH. Metrological traceability in clinical biochemistry. Ann Clin Biochem. 2011;48(Pt 5):393–409.
44. Home – JCTLM. Available from: http://www.jctlm.org/. Accessed Aug 10th, 2017.
45. Armbruster D, Miller RR. The Joint Committee for Traceability in Laboratory Medicine (JCTLM): a global approach to promote the standardisation of clinical laboratory test results. Clin Biochem Rev. 2007;28(3):105–13.
46. Jones GRD, Jackson C. The Joint Committee for Traceability in Laboratory Medicine (JCTLM) - its history and operation. Clin Chim Acta Int J Clin Chem. 2016;453:86–94.
47. Panteghini M, Ceriotti F. Obtaining reference intervals traceable to reference measurement systems: is it possible, who is responsible, what is the strategy? Clin Chem Lab Med. 2011;50(5):813–7.
48. Bais R, Armbruster D, Jansen RTP, Klee G, Panteghini M, Passarelli J, et al. Defining acceptable limits for the metrological traceability of specific measurands. Clin Chem Lab Med. 2013;51(5):973–9.
49. Stavelin A, Albe X, Meijer P, Sarkany E, MacKenzie F. An overview of the European Organization for External Quality Assurance Providers in Laboratory Medicine (EQALM). Biochem Med. 2017;271(1):30–6.
50. International Organisation for Standardisation 13528. Statistical methods for use in proficiency testing by interlaboratory comparison. Geneva, Switzerland, 2015.
51. Coucke W, Soumali MR. Demystifying EQA statistics and reports. Biochem Med. 2017;27(1):37–48.
52. Miller WG, Jones GRD, Horowitz GL, Weykamp C. Proficiency Testing/External Quality Assessment: Current Challenges and Future Directions. Clin Chem. 2011;57(12):1670–80.
53. Jones GRD, Albarede S, Kesseler D, MacKenzie F, Mammen J, Pedersen M, et al. Analytical performance specifications for external quality assessment – definitions and descriptions. Clin Chem Lab Med. 2017;55(7):949–55.
54. Sciacovelli L, Secchiero S, Zardo L, Plebani M. External Quality Assessment Schemes: need for recognised requirements. Clin Chim Acta Int J Clin Chem. 2001;309(2):183–99.
55. Dallas JGR. Analytical performance specifications for EQA schemes – need for harmonisation. Clin Chem Lab Med. 2015;53(6):919–924.
56. Sandberg S, Fraser CG, Horvath AR, Jansen R, Jones G, Oosterhuis W, et al. Defining analytical performance specifications: Consensus Statement from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine. Clin Chem Lab Med. 2015;53(6):833–5.
57. C. G. Fraser AK D Kenny, P Hyltoft Petersen. Strategies to set global quality specifications in laboratory medicine. Scand J Clin Lab Invest. 1999;59(7):477–8.

58. Horvath AR, Bossuyt PMM, Sandberg S, John AS, Monaghan PJ, Verhagen-Kamerbeek WDJ, et al. Setting analytical performance specifications based on outcome studies – is it possible? Clin Chem Lab Med. 2015;53(6):841–8.
59. Perich C, Minchinela J, Ricós C, Fernández-Calle P, Alvarez V, Doménech MV, et al. Biological variation database: structure and criteria used for generation and update. Clin Chem Lab Med. 2014;53(2):299–305.
60. Desirable Biological Variation Database specifications – Westgard. Available from: https://www.westgard.com/biodatabase1.htm. Accessed Dec 20th, 2017.
61. Jones GRD. Laboratory analytical quality – the process continues. Clin Chem Lab Med. 2016;54(8):1275–6.
62. Ceriotti F, Fernandez-Calle P, Klee GG, Nordin G, Sandberg S, Streichert T, et al. Criteria for assigning laboratory measurands to models for analytical performance specifications defined in the 1st EFLM Strategic Conference. Clin Chem Lab Med. 2016;55(2):189–194.
63. Coucke W, China B, Delattre I, Lenga Y, Van Blerk M, Van Campenhout C, et al. Comparison of different approaches to evaluate External Quality Assessment Data. Clin Chim Acta. 2012;413(5–6):582–6.
64. Bais R. EQA from an Australian Perspective. Clin Biochem Rev. 2007;28(4):175–8.
65. Labquality EQAS. Available from: https://www.labquality.fi/en/external-quality-assessment/. Accessed Nov 4th, 2017.
66. Miller WG, Myers GL. Commutability Still Matters. Clin Chem. 2013;59(9):1291–3.
67. Vlašić Tanasković J, Coucke W, Leniček Krleža J, Vuković Rodriguez J. Peer groups splitting in Croatian EQA scheme: a trade-off between homogeneity and sample size number. Clin Chem Lab Med. 2017;55(4):539–45.
68. Thienpont LM, Stöckl D, Friedecký B, Kratochvíla J, Budina M. Trueness verification in European external quality assessment schemes: time to care about the quality of the samples. Scand J Clin Lab Invest. 2003;63(3):195–201.
69. Miller WG. Specimen materials, target values and commutability for external quality assessment (proficiency testing) schemes. Clin Chim Acta. 2003;327(1):25–37.
70. Ferrero CA, Carobene A, Ceriotti F, Modenese A, Arcelloni C. Behavior of frozen serum pools and lyophilized sera in an external quality-assessment scheme. Clin Chem. 1995;41(4):575–580.
71. Cobbaert C, Weykamp C, Baadenhuijsen H, Kuypers A, Lindemans J, Jansen R. Selection, Preparation, and Characterization of Commutable Frozen Human Serum Pools as Potential Secondary Reference Materials for Lipid and Apolipoprotein Measurements: Study within the Framework of the Dutch Project "Calibration 2000." Clin Chem. 2002;48(9):1526–38.
72. Cattozzo G, Franzini C, d'Eril GM. Commutability of Calibration and Control Materials for Serum Lipase. Clin Chem. 2001;47(12):2108–13.
73. Thienpont LM, Uytfanghe KV, Marriott J, Stokes P, Siekmann L, Kessler A, et al. Feasibility Study of the Use of Frozen Human Sera in Split-Sample Comparison of Immunoassays with Candidate Reference Measurement Procedures for Total Thyroxine and Total Triiodothyronine Measurements. Clin Chem. 2005;51(12):2303–11.
74. Henriksen GM, Pedersen MM, Nørgaard I, Blom M, Blou L, Blaabjerg O, et al. Minimally processed fresh frozen human reference sera: preparation, testing, and application to international external quality assurance. Scand J Clin Lab Invest. 2004;64(4):293–308.
75. Emons H. The 'RM family'—Identification of all of its members. Accreditation Qual Assur. 2006;10(12):690–1.

76. Vesper HW, Miller WG, Myers GL. Reference Materials and Commutability. Clin Biochem Rev. 2007;28(4):139–47.
77. Rej R, Jenny RW, Bretaudiere J-P. Quality control in clinical chemistry: characterization of reference materials. Talanta. 1984;31(10, Part 2):851–62.
78. Franzini C. Commutability of reference materials in clinical chemistry. J Int Fed Clin Chem. 1993;5(4):169–73.
79. Joint Committee for Guides in Metrology (JCGM). The International vocabulary of metrology - Basic and general concepts and associated terms (VIM), 3rd edn. JCGM 200:2012.
80. Schimmel H, Zegers I, Emons H. Standardization of protein biomarker measurements: Is it feasible? Scand J Clin Lab Invest. 2010;70(sup242):27–33.
81. Zegers I, Beetham R, Keller T, Sheldon J, Bullock D, MacKenzie F, et al. The Importance of Commutability of Reference Materials Used as Calibrators: The Example of Ceruloplasmin. Clin Chem. 2013;59(9):1322–9.
82. Miller WG, Erek A, Cunningham TD, Oladipo O, Scott MG, Johnson RE. Commutability Limitations Influence Quality Control Results with Different Reagent Lots. Clin Chem. 2011;57(1):76–83.
83. Clinical and Laboratory Standards Institute. Evaluation of matrix effects; Approved Guideline – Second edition. CLSI document EP14-A2, Wayne, PA USA: Clinical and Laboratory Standards Institute; 2005.
84. Clinical and Laboratory Standards Institute. Interference Testing in Clinical Chemistry; Approved Guideline - Second Edition. CLSI document EP07-A2, Wayne,PA: Clinical and Laboratory Standards Institute; 2005.
85. Li Y, Wang J, Huang X, Zeng R, Zhang Q, Lin H, et al. Matrix Effects in Proficiency Testing Materials Influence the Accurate Measurement of Gamma-Glutamyltransferase Activity. Clin Lab. 2016;62(10):1941–5.
86. Meng Q, Zhou W, Zhang C, Zeng J, Zhao H, Zhang T, et al. Serum triglyceride measurements: the commutability of reference materials and the accuracy of results. Clin Chem Lab Med. 2017;55(9):1284–90.
87. Houcke SKV, Rustad P, Stepman HCM, Kristensen GBB, Stöckl D, Røraas TH, et al. Calcium, Magnesium, Albumin, and Total Protein Measurement in Serum as Assessed with 20 Fresh-Frozen Single-Donation Sera. Clin Chem. 2012;58(11):1597–9.
88. Braga F, Frusciante E, Infusino I, Aloisio E, Guerra E, Ceriotti F, et al. Evaluation of the trueness of serum alkaline phosphatase measurement in a group of Italian laboratories. Clin Chem Lab Med. 2016;55(3):e47–50.
89. Lenicek Krleza J, Celap I, Vlasic Tanaskovic J. External Quality Assessment in Croatia: problems, challenges, and specific circumstances. Biochem Med. 2017;27(1):86–92.
90. Jansen R, Jassam N, Thomas A, Perich C, Fernandez-Calle P, Faria AP, et al. A category 1 EQA scheme for comparison of laboratory performance and method performance: An international pilot study in the framework of the Calibration 2000 project. Clin Chim Acta. 2014;432:90–8.
91. Bretaudiere JP, Dumont G, Rej R, Bailly M. Suitability of control materials. General principles and methods of investigation. Clin Chem. 1981 Jun 1;27(6):798–805.
92. Bretaudiere J-P, Rej R, Drake P, Vassault A, Bailly M. Suitability of control materials for determination of alpha-amylase activity. Clin Chem. 1981;27(6):806–815.
93. Kimberly MM, Vesper HW, Caudill SP, Cooper GR, Rifai N, Dati F, et al. Standardization of Immunoassays for Measurement of High-Sensitivity C-reactive Protein. Phase I: Evaluation of Secondary Reference Materials. Clin Chem. 2003;49(4):611–6.

94. Eckfeldt J, Copeland K. Accuracy verification and identification of matrix effects. Arch Pathol Lab Med. 1993(117):381–6.
95. Clinical and Laboratory Standards Institute. Evaluation of Commutability of Processed Samples; Approved Guideline-Third Edition. CLSI document EP14-A3. Wayne,PA: Clinical and Laboratory Standards Institute; 2014.
96. Franzini C, Ceriotti F. Impact of reference materials on accuracy in clinical chemistry. Clin Biochem. 1998;31(6):449–57.
97. Cattozzo G, Guerra E, Ceriotti F, Franzini C. Commutable Calibrator with Value Assigned by the IFCC Reference Procedure to Harmonize Serum Lactate Dehydrogenase Activity Results Measured by 2 Different Methods. Clin Chem. 2008;54(8):1349–55.
98. van den Besselaar AMHP, van Rijn CJJ, Cobbaert CM, Reijnierse GLA, Hollestelle MJ, Niessen RWLM, et al. Fibrinogen determination according to Clauss: commutability assessment of International and commercial standards and quality control samples. Clin Chem Lab Med. 2017;55(11):1761–1769.
99. Baadenhuijsen H, Steigstra H, Cobbaert C, Kuypers A, Weykamp C, Jansen R. Commutability assessment of potential reference materials using a multicenter split-patient-sample between-field-methods (twin-study) design: study within the framework of the Dutch project "Calibration 2000." Clin Chem. 2002;48(9):1520–1525.
100. Passing H, Bablok null. A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, Part I. J Clin Chem Clin Biochem Z Klin Chem Klin Biochem. 1983;21(11):709–20.
101. Bilic-Zulle L. Comparison of methods: Passing and Bablok regression. Biochem Med. 2011;21(1):49–52.
102. Ricós C, Juvany R, Jiménez CV, Perich C, Minchinela J, Hernández A, et al. Procedure for studying commutability validated by biological variation. Clin Chim Acta Int J Clin Chem. 1997;268(1–2):73–83.
103. Korzun WJ, Nilsson G, Bachmann LM, Myers GL, Sakurabayashi I, Nakajima K, et al. Difference in Bias Approach for Commutability Assessment: Application to Frozen Pools of Human Serum Measured by 8 Direct Methods for HDL and LDL Cholesterol. Clin Chem. 2015;61(8):1107–13.
104. Delatour V, Liu Q, Vesper HW. Commutability Assessment of External Quality Assessment Materials with the Difference in Bias Approach: Are Acceptance Criteria Based on Medical Requirements too Strict? Clin Chem. 2016;62(12):1670–1671.
105. Miller WG, Schimmel H, Rej R, Greenberg N, Ceriotti F, Burns C, et al. IFCC Working Group Recommendations for Assessing Commutability Part 1: General Experimental Design. Clin Chem. 2018;64(3):447–54.
106. Nilsson G, Budd JR, Greenberg N, Delatour V, Rej R, Panteghini M, et al. IFCC Working Group Recommendations for Assessing Commutability Part 2: Using the Difference in Bias between a Reference Material and Clinical Samples. Clin Chem. 2018;64(3):455–64.
107. Budd JR, Weykamp C, Rej R, MacKenzie F, Ceriotti F, Greenberg N, et al. IFCC Working Group Recommendations for Assessing Commutability Part 3: Using the Calibration Effectiveness of a Reference Material. Clin Chem. 2018;64(3):465–74.
108. Zakon o krvi i krvnim pripravcima. Available from: https://www.zakon.hr/z/511/Zakon-o-krvi-i-krvnim-pripravcima (In Croatian). Accessed March 21st, 2019.
109. Grubbs FE. Procedures for Detecting Outlying Observations in Samples. Technometrics. 1969;11(1):1–21.
110. Hsu JC. Multiple Comparisons: Theory and Methods. CRC Press, Boca Raton, FL, USA; 1996.

111. CROQALM Analytical Performance Specifications. 2017. Available from: https://croqalm.hdmblm.hr/images/tablica/DOD_2017_MODULI_za_web_ver2.pdf (In Croatian). Accessed March 21st, 2019.
112. C. Ricós, V. Alvarez, F. Cava, J. V. Current databases on biological variation: pros, cons and progress. Scand J Clin Lab Invest. 1999 Jan;59(7):491–500
113. Efron B, Tibshirani RJ. An Introduction to the Bootstrap. Boca Raton, FL, USA: CRC Press; 1994.

8. ABBREVIATIONS

TTT – Total Testing Process

CLIA Clinical Laboratory Improvement Amendments

RiliBÄK Guideline of the German Medical Association on Quality Assurance in Medical Laboratory Examinations

CCMB Croatian Chamber of Medical Biochemists (CCMB)

IQC Internal Quality Control

EQA External Quality Assessment

MP Measurement procedure

MBL Medical Biochemical Laboratories

PT Proficiency Testing

KDIGO Kidney Disease Improving Global Outcomes

APS Analytical Performance Specifications

SI Système International

BIPM *Bureau International des Poids et Mesures*

IFCC International Federation of Clinical Chemistry and Laboratory Medicine

ILAC International Laboratory Accreditation Cooperation

APS Analytical Performance Specifications

EFLM European Federation of Clinical Chemistry and Laboratory Medicine

TFG-APSEQA Task and Finish Group on Performance Specifications for EQA

VIM International Vocabulary of Metrology

CRP C-reactive protein

CLSI Clinical and Laboratory Standards Institute

CDC Centers for Disease Control

IFCC-WGC – IFCC Working Group on Commutability

9. APPENDIX

MP pairs excluded from commutability evaluation in the EQA survey 2

MP 1	MP 2	Mean MP1 (serum)	Mean MP2 (serum)	Mean MP1 (control)	Mean MP2 (control)	% harmonisation	% commutability
ALT							
IFCC- BC AU	IFCC- SD	7.28	13.25	6.79	11.42	0.1	67.5
IFCC- SD	Photometry UV- AA	13.25	7	11.42	6.56	2.5	66.9
IFCC- SD	Photometry UV- BC AU	13.25	6.64	11.42	6.34	0	58.8
IFCC- SD	Photometry UV- RCc	13.25	6.62	11.42	6.25	2.8	43.9
IFCC- SD	Photometry UV- RCI	13.25	6.43	11.42	6.07	0	79.7
IFCC- SD	Photometry UV- RH	13.25	6.89	11.42	6.21	0	79.1
AMY							
IFCC- BC AU	IFCC- RH	55.58	52.36	51.99	47.35	100	61.4
CALCIUM							
Arsenaso III- AA	cresolphthalein- SD	2.42	2.33	2.25	2.18	51.6	85
Arsenaso III- BC AU	cresolphthalein- BC AU	2.41	2.38	2.27	2.23	79.3	83.5
Arsenaso III- BC AU	cresolphthalein- SD	2.41	2.33	2.27	2.18	32.7	66.5
CHLORIDE							
Indirect ISE- AA	Indirect ISE- BC AU	105.27	105.02	121.1	135.18	100	0
Indirect ISE- BC AU	Indirect ISE- SD	105.02	102.57	135.18	122	91.5	0
CREATININE							
Compensated Jaffe- AA	Compensated Jaffe- BC AU	66	60.33	66.71	56.91	94.5	5.3
Compensated Jaffe- AA	Compensated Jaffe- RCc	66	64.42	66.71	59.83	100	66.2
Compensated Jaffe- AA	Compensated Jaffe- RCI	66	66.46	66.71	61	100	76.9
Compensated Jaffe- AA	Compensated Jaffe- RH	66	63.42	66.71	62.69	100	89.5
Compensated Jaffe- AA	Enzymatic method- BC AU	66	59.8	66.71	56.33	100	35.2
Compensated Jaffe- BC AU	Compensated Jaffe- RCc	60.33	64.42	56.91	59.83	81.5	91.6
Compensated Jaffe- BC AU	Compensated Jaffe- RCI	60.33	66.46	56.91	61	49	90.5
Compensated Jaffe- BC AU	Compensated Jaffe- RH	60.33	63.42	56.91	62.69	100	39.6

Compensated Jaffe- BC AU	Compensated Jaffe- SD	60.33	63	56.91	63.22	100	90.9
Compensated Jaffe- BC AU	Non-compensated Jaffe- BC AU	60.33	72.55	56.91	69.86	0	86.8
Compensated Jaffe- RCc	Non-compensated Jaffe- BC AU	64.42	72.55	59.83	69.86	20.3	74.3
Compensated Jaffe- RCI	Non-compensated Jaffe- BC AU	66.46	72.55	61	69.86	78.1	76.9
Compensated Jaffe- RH	Enzymatic method- BC AU	63.42	59.8	62.69	56.33	100	91.1
Compensated Jaffe- RH	Non-compensated Jaffe- BC AU	63.42	72.55	62.69	69.86	47	91.1
Compensated Jaffe- SD	Non-compensated Jaffe - BC AU	63	72.55	63.22	69.86	29.1	48.1
Enzymatic method- BC AU	Non-compensated Jaffe- BC AU	59.8	72.55	56.33	69.86	0	62.7
HDL							
Homogenous - BC AU	Homogenous- SD	1.61	1.7	1.5	1.6	65.4	77.7
LDH							
IFCC- BC AU	IFCC- SD	133.73	126	121.54	111.33	87	74
IFCC- RCI	IFCC- SD	135.57	126	123.54	111.33	91.9	85.7
PHOSPHATE							
Ammonium-molybdate- AA	Ammonium-molybdate- BC AU	0.8	0.79	0.76	0.73	100	84.9
BILIRUBIN							
Diazo- AA	Diazo- BC AU	6.6	7.37	39.27	47.79	100	0
Diazo- AA	Diazo- HP	6.6	6.5	39.27	45.17	100	62.6
Diazo- AA	Diazo- RH	6.6	6.45	39.27	42.48	100	83.2
Diazo- AA	Diazo- SD	6.6	6.08	39.27	43.15	100	84.4
Diazo- BC AU	Diazo- RCc	7.37	5.15	47.79	39.77	100	0
Diazo- BC AU	Diazo- RCI	7.37	5.19	47.79	38.93	88.7	0
Diazo- BC AU	Diazo- SD	7.37	6.08	47.79	43.15	100	87.7
Diazo- HP	Diazo- RCc	6.5	5.15	45.17	39.77	100	70.2
Diazo- HP	Diazo- RCI	6.5	5.19	45.17	38.93	100	30.2
PROTEINS							
Biuret- AA	Biuret- BC AU	66.2	68.16	61.1	63.49	98.9	85.5
URATE							
Uricase,POD - AA	Uricase- SD	176.83	157.3	164	149.33	59	82.2
Uricase,POD - BC AU	Uricase- SD	173.11	157.3	159.6	149.33	81.6	83.2

UREA							
Urease,GLDH- BC AU	Urease,GLDH-RCI	4.47	4.22	14.77	14.36	78.4	83.2
Urease,GLDH- BC AU	Urease,GLDH-SD	4.47	4.6	14.77	15.01	86.7	89.1
Urease,GLDH- RCI	Urease,GLDH-RH	4.22	4.57	14.36	14.9	79.3	85.9
Urease,GLDH- RCI	Urease,GLDH-SD	4.22	4.6	14.36	15.01	45	85.10

10. BIOGRAPHY

Jelena Vlašić Tanasković was born on 20th of September in 1972 in Split, Croatia. She graduated from El Cerrito High School, CA, USA in 1991. In 1999 she started studying at Faculty of Pharmacy and Biochemistry, University of Zagreb. After graduation in 1998, she received the grant from Scientific Committee of 17th IFCC Congress for poster presentation "CAG repeat analysis by Expand Long PCR in molecular diagnosis of Huntington Disease".

Since 1999 Jelena works at Department of Laboratory Diagnostics, General Hospital Pula. In 2003 she started her residency at Clinical Department of Laboratory Diagnostics, Clinical Hospital Zagreb and specialised in Medical Biochemistry in 2008.

From 2013 she works at the Croatian Centre for Quality Assessment in Laboratory Medicine (CROQALM), Croatian Society of Medical Biochemistry and Laboratory Medicine as a member of the professional board and a chair of biochemistry module in CROQALM.

She co-authored 8 peer-reviewed journal articles, one book chapter and participated in many international scientific meetings and congresses with poster presentations.

Jelena is a member of Croatian Chamber of Medical Biochemists and Croatian Society of Medical Biochemistry and Laboratory Medicine.

List of journal articles

1. Vuljanić D, Dojder A, Špoljarić V, Saračević A, Dukić L, Leniček Krleža J, Vlašić Tanasković J, Maradin I, Grzunov A, Vogrinc Ž, Šimundić AM. Analytical verification of 12 most commonly used urine dipsticks in Croatia: comparability, repeatability and accuracy. Biochem Med.2019;29(1):010708.

2. Coucke W, Vlašić Tanasković J, Bouacida L, Broeders S, China B, Demarteau M, Ghislain V, Lenga Y, Van Blerk M, Vandevelde N, Verbeke H, Wathlet S, Soumali M. Alternative sample homogeneity test for quantitative and qualitative Proficiency Testing schemes. Anal Chem.2019;91(3):1847-1854.

3. Leniček Krleža J, Čelap I, Vlašić Tanaskovic J. External Quality Assessment in Croatia: problems, challenges, and specific circumstances. Biochem Med. 2017;27(1):86-92.

4. Vlašić Tanasković J, Coucke W, Leniček Krleža J, Vuković Rodriguez J. Peer groups splitting in Croatian EQA scheme: a trade-off between homogeneity and sample size number. Clin Chem Lab Med 2017;55(4):539-545.

5. Županić, D., Vlašić-Tanasković, J., Šmalcelj, R., Kes, P., Kušec, V. Bone markers in metabolic bone disorder in patients on chronic hemo dialysis and kidney transplant recipients. Biochem Med 2006;16(2):137-149.

6. Jelena Vlašić Tanasković, Jadranka Sertić. Dynamic mutations in human genome: a review of triplet repeat diseases. Biochem Med 2004,14(3-4):101-108.

7. Hećimović S, Klepac N, Vlašić J, Vojta A, Janko D, Skarpa-Prpić I, Canki-Klain N, Marković D, Bozikov J, Relja M, Pavelić K.Genetic background of Huntington disease in Croatia: Molecular analysis of CAG, CCG, and Delta2642 (E2642del) polymorphisms.Hum Mutat. 2002 Sep;20(3):233.

8. Hećimović S, Vlasić J, Barisić L, Marković D, Culić V, Pavelić K. A simple and rapid analysis of triplet repeat diseases by expand long PCR.Clin Chem Lab Med. 2001;39(12):1259-62.

Book chapter

Jadranka Sertić, Jelena Vlašić Tanasković. Molekularna analitika nasljednih neuroloških bolesti. In: Barišić N. eds. Pedijatrijska neurologija. Zagreb: Medicinska naklada; 2009.

BASIC DOCUMENTATION CARD

University of Zagreb
Faculty of Pharmacy and Biochemistry Doctoral Thesis

Commutability evaluation of control samples within external quality assessment programs of medical biochemical laboratories

Jelena Vlašić Tanasković

Introduction and aim: External quality assessment (EQA) is an integral part of quality management systems in medical biochemical laboratories enabling monitoring of individual results as well as harmonisation and standardisation of measurement procedures (MPs) used in the clinical setting. Commutability of control samples is a major prerequisite for assessing laboratory and MP performance according to the unique target value. Commutable samples show the same properties in different MPs as well as patient samples. Commutability is usually evaluated using regression analysis and statistically determined criteria of acceptance without taking into consideration analytical performance specifications for the analyte. The aim of this research is to propose a new model for the evaluation of commutability criteria using analytical performance specifications for each analyte within the EQA program for medical biochemical laboratories.
Materials and methods: Lyophilised control samples were distributed together with native and spiked serum samples to all participants of Croatian EQA (CROQALM). The participants analysed both samples using routine MPs. Commutability of control samples was evaluated using the results of two kinds of samples and newly proposed false flagging method. The results for commutability were compared to statistically determined commutability criteria obtained by recommended regression analysis for commutability evaluation of EQA control samples.
Results: Three lyophilised EQA control samples were evaluated for commutability for 22 biochemistry analytes and related MPs used in medical biochemical laboratories. The controls were found commutable for 13 analytes: AMY, AST, CK, glucose, iron, LDH, phosphate, potassium, sodium, proteins, triglycerides, urate and urea. High noncommutability of control materials was found for chloride in all three control samples and HDL-cholesterol, AP, creatinine and calcium in two out of three control samples. Unequal criteria in statistically defined commutability limits resulted in commutability conclusions that are dependent on measurement results of patient serum samples by evaluated MPs.
Conclusions: The false flagging method, proposed in this thesis, can be used for evaluating commutability of control samples within the EQA program of medical biochemical laboratories. The commutability limits are equally designed for all MP combinations and connected to established analytical performance specifications of the analytes.

Number of pages:142; number of figures:15; number of tables:21, original in English language

Keywords: commutability, external quality assessment, false flagging method

Supervisors: Wim Coucke, Ph.D.
Assoc. Prof. Jadranka Vuković Rodrìguez
Reviewers: Assoc. Prof. Dunja Rogić
Prof. József Petrik
Assoc. Prof. Ana-Maria Šimundić
Accepted: 12th June 2019

TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište u Zagrebu
Farmaceutsko-biokemijski fakultet Doktorski rad

Prosudba komutabilnosti kontrolnih uzoraka u programima vanjske procjene kvalitete medicinsko-biokemijskih laboratorija

Jelena Vlašić Tanasković

Uvod i cilj: Vanjska procjena kvalitete sastavni je dio sustava za upravljanje kvalitetom medicinsko-biokemijskih laboratorija. Osim prosudbe mjernih rezultata, vanjska procjena kvalitete ima za svrhu praćenje globalnih ciljeva harmonizacije i standardizacije mjernih postupaka (MP) koji se koriste u laboratorijima. Komutabilnost kontrolnih uzoraka nužan je preduvjet za valjanu prosudbu kvalitete prema jedinstvenoj ciljnoj vrijednosti. Komutabilni uzorci pokazuju jednaka svojstva kao i uzorci pacijenata u različitim MP. Komutabilnost se uobičajeno procjenjuje korištenjem regresijske analize i statističkih kriterija za prosudbu, ne uzimajući u obzir analitičke ciljeve kvalitete ispitivanog analita. Stoga je cilj ovog doktorskog rada postavljanje i validacija nove metode za prosudbu komutabilnosti kontrolnih uzoraka uzimajući u obzir postavljene analitičke ciljeve kvalitete u sklopu vanjske procjene kvalitete medicinsko-biokemijskih laboratorija .
Materijali i metode. Medicinsko-biokemijskim laboratorijima koji sudjeluju u vanjskoj procjeni kvalitete CROQALM poslani su liofilizirani komercijalni kontrolni uzorci zajedno sa svježim uzorcima seruma, te svježim uzorcima seruma uz dodatak glukoze, ureje, natrija, kalija, klorida i bilirubina. Uzorci su analizirani upotrebom standardnih MP, a komutabilnost komercijalnih kontrolnih uzoraka ispitana je korištenjem dvije vrste uzoraka i novom predloženom 'metodom lažnog odstupanja' (engl. *false flagging method*). Dobiveni rezultati uspoređivani su sa statističkim kriterijima prosudbe komutabilnosti kontrolnih uzoraka u okviru vanjske procjene kvalitete medicinsko-biokemijskih laboratorija.
Rezultati: Metodom lažnog odstupanja ispitana je komutabilnost kontrolnih uzoraka za 22 analita i 331-426 parova MP koji se koriste u rutinskom radu laboratorija. Sva tri kontrolna uzorka pokazuju komutabilnost za većinu kombinacija MP za mjerenje amilaza, AST, CK, glukoze, željeza, LDH, fosfata, kalija, natrija, proteina, triglicerida, urata i ureje. Nekomutabilnost sva tri kontrolna uzorka dokazana je za kloride, te HDL-kolesterol, AP, kreatinin i kalcij u dvije kontrole. Neujednačenost statistički postavljenih kriterija za prosudbu komutabilnosti rezultira zaključcima koja uvelike zavise o rezultatima mjerenja uzoraka seruma pacijenata na ispitivanim MP.
Zaključci: Metoda lažnog odstupanja predložena u ovom radu predstavlja novi pristup u prosudbi komutabilnosti i može se primijeniti za veliki broj analita i MP u okviru vanjske procjene kvalitete medicinsko-biokemijskih laboratorija. Pri tome su kriteriji prosudbe jednoznačni za sve parove MP, omogućavajući prosudbu kliničke i/ili analitičke jednakovrijednosti kontrolnih uzoraka prema dijagnostičkim značajkama samog analita.

Broj stranica:142; broj slika: 15; broj tablica: 21; izvornik je na engleskom jeziku.

Ključne riječi: komutabilnost, vanjska procjena kvalitete, metoda lažnog odstupanja

Mentori: dr. sc. Wim Coucke
izv. prof. dr. sc. Jadranka Vuković Rodrìguez

Ocjenjivači: izv. prof. dr. sc. Dunja Rogić
prof. dr. sc. József Petrik
nasl. izv. prof. dr. sc. Ana-Maria Šimundić

Datum prihvaćanja rada: 12. lipnja 2019. godine

Jelena Vlašić Tanasković*, Wim Coucke, Jasna Leniček Krleža and Jadranka Vuković Rodriguez

Peer groups splitting in Croatian EQA scheme: a trade-off between homogeneity and sample size number

DOI 10.1515/cclm-2016-0284
Received April 8, 2016; accepted July 29, 2016

Abstract

Background: Laboratory evaluation through external quality assessment (EQA) schemes is often performed as 'peer group' comparison under the assumption that matrix effects influence the comparisons between results of different methods, for analytes where no commutable materials with reference value assignment are available. With EQA schemes that are not large but have many available instruments and reagent options for same analyte, homogenous peer groups must be created with adequate number of results to enable satisfactory statistical evaluation. We proposed a multivariate analysis of variance (MANOVA)-based test to evaluate heterogeneity of peer groups within the Croatian EQA biochemistry scheme and identify groups where further splitting might improve laboratory evaluation.

Methods: EQA biochemistry results were divided according to instruments used per analyte and the MANOVA test was used to verify statistically significant differences between subgroups. The number of samples was determined by sample size calculation ensuring a power of 90% and allowing the false flagging rate to increase not more than 5%. When statistically significant differences between subgroups were found, clear improvement of laboratory evaluation was assessed before splitting groups.

Results: After evaluating 29 peer groups, we found strong evidence for further splitting of six groups. Overall improvement of 6% reported results were observed, with the percentage being as high as 27.4% for one particular method.

Conclusions: Defining maximal allowable differences between subgroups based on flagging rate change, followed by sample size planning and MANOVA, identifies heterogeneous peer groups where further splitting improves laboratory evaluation and enables continuous monitoring for peer group heterogeneity within EQA schemes.

Keywords: external quality assessment; multivariate analysis of variance; peer group.

Introduction

External quality assessment (EQA) provides an essential tool for medical laboratories for evaluating assay performances and establishing quality standards. One of the key purposes of such an evaluation is standardization of analytical measurement procedures which, in turn, would yield comparable patient results across a variety of measurement procedures and calibration details in different laboratories. Depending on the EQA scheme design and type of control material used, results are usually validated based on distance of results from the target value [1]. EQA schemes distributing samples with target values assessed by reference methods are now recognized as category 1 or 2 schemes [2] being able to assess individual laboratory performance and monitor standardization and traceability of laboratory methods used. Target value assessment by reference method depends on the commutability of control material and the availability of a reference method for a particular analyte. When the commutability of a sample is unknown, a reference measurement value cannot be used as target value since it is not possible to determine whether the observed difference from the target is caused

***Corresponding author: Jelena Vlašić Tanasković,** Department of Laboratory Diagnostics, General Hospital Pula, Zagrebačka 30, 52100 Pula, Croatia, Phone: +385 52 376 811, E-mail: jelena.vlasictanaskovic@gmail.com; and Croatian Centre for Quality Assessment in Laboratory Medicine (CROQALM), Croatian Society of Medical Biochemistry and Laboratory Medicine, Zagreb, Croatia
Wim Coucke: Quality of Medical Laboratories, Scientific Institute of Public Health, Brussels, Belgium
Jasna Leniček Krleža: Croatian Centre for Quality Assessment in Laboratory Medicine (CROQALM), Croatian Society of Medical Biochemistry and Laboratory Medicine, Zagreb, Croatia; and Department of Laboratory Diagnostics, Children's Hospital Zagreb, Zagreb, Croatia
Jadranka Vuković Rodriguez: Faculty of Pharmacy and Biochemistry, University of Zagreb, Zagreb, Croatia

by calibration bias or assays not traceable to higher order reference methods or matrix-related bias of unknown magnitude. Although EQA organizers should strive to use commutable material [3], commutability of EQA material is not always assured due to processing steps used to enhance sample stability or allow distribution on a large scale [4]. Samples used in EQA schemes are often considered commutable based on stringency of their preparation, but commutability should not always be assumed a priori in highly processed materials without a validation for all combination of measurement methods used. In case commutability is not assured, results are usually categorized into peer groups within which results are obtained by a similar technology. Target values for peer groups are generally calculated for each peer group individually and are called 'consensus values'. Members of peer groups are expected to have the same matrix-related bias for a given EQA sample [2].

Whenever statistical techniques are used to calculate consensus values and deviation of results from consensus values, two antagonistic criteria play a role. Peer groups should be as large as possible to produce reliable statistics, and they should be as homogenous as possible to produce correct statistical calculations. Hence, the creation of peer groups is a compromise between obtaining satisfactory homogeneity within the peer group, and maintaining satisfactory peer group sizes. A test is proposed to identify heterogeneity of peer groups and has been applied to the Croatian EQA scheme for general biochemistry. It helps identifying heterogeneity of peer groups that consist of different subgroups of data that were obtained on analyzers that use the same analytical principle, but are from different manufacturers.

Croatian laboratories often use 'open systems', in which reagents and instruments may be from different manufacturers. Due to the large number of possible combinations of instruments and reagents, the EQA organizer initially identifies peer groups based on the method or analytical principle, without making any distinction between equipment, reagents or calibrators used. Although this approach is useful in addressing harmonization of individual laboratories' results, inadequate standardization within the same method [5] and the influence of dominant instruments on the consensus value are evident shortcomings of method-based peer groups. The aim of this study was to identify peer groups for their homogeneity and to verify that splitting based on equipment manufacturer would improve laboratory evaluation. Currently, laboratories are evaluated based on their percentage difference from their peer group mean with predefined analyte-specific allowable limits of performance (ALP), based on biological variation, statistical analysis, expert opinion and combination of approaches [6].

Materials and methods

Study design

The results from four rounds from the Croatian EQA biochemistry scheme in 2014 and 2015 were analyzed. The analytes included enzymes [aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase (AP), lipase], electrolytes (sodium, potassium, chloride, calcium, phosphorous, magnesium), substrates (glucose, total bilirubin, creatinine, urea, uric acid, total iron binding capacity), proteins (total protein, albumin, CRP) and total cholesterol. The data submitted by laboratories were first divided into peer groups according to the analyte and analytical method and peer groups were divided into subgroups according to the manufacturer. Every subgroup that contained at least 10 laboratories for every sample was considered as an independent subgroup. Subgroups of smaller size were all joined together into a subgroup called 'Other'. The data were graphically inspected using two scatter plots containing data from two rounds each. The rounds on the scatter plots were combined to assess, when possible, both lower and higher concentration levels on the same graph.

Statistical analysis

In a first instance, the effect of peer group heterogeneity on the laboratory evaluation was modeled. The assumption was made that reported results follow a Gaussian distribution and that the mean and standard deviation are not influenced by outliers. The last criterion can be achieved by removing outliers before calculating mean and standard deviation or by using robust estimates of the mean and standard deviation.

When a peer group is homogeneous, any subgroup consisting of results that have been obtained by analyzers from the same manufacturer, has the same mean. The chance of falsely flagging a result is given by:

$$\text{false flagging} = 2 * \mathrm{F}(m_{\mathrm{hom}}(1-d(\%)/100),\ m_{\mathrm{hom}},\ \mathrm{s}_{\mathrm{hom}}), \tag{1}$$

with F(x, m, s) being the cumulative probability function of a Gaussian distribution with mean m and standard deviations. The parameters m_{hom} and s_{hom} are the mean and standard deviation of the homogeneous peer group and d(%) is the deviation from the assigned value, defined by the ALP, expressed as a percentage. This probability increases with decreasing d and increasing standard deviation.

When a peer group is heterogeneous, the mean and standard deviation of one or more peer subgroups may be different from the mean and standard deviation of the whole group. For example, when there are exactly two subgroups, with unequal means x_1 and x_2, respectively, with $x_2 = x_1 + \delta$, the mean of the heterogeneous peer group m_{het} is given by:

$$m_{\mathrm{het}} = x_1 + p * \delta, \tag{2}$$

where p is the fraction of data belonging to the group with mean x_2.

- 10.1515/cclm-2016-0284

If the two subgroups have the same standard deviation s_1, the standard deviation of the heterogeneous group s_{het} becomes:

$$s_{het} = \sqrt{\frac{(n-2)s_1^2 + pn\delta^2 - p^2 n\delta^2}{n-1}}. \quad (3)$$

When n is large, this formula can be simplified to:

$$s_{het} = \sqrt{s_1^2 + p\delta^2 - p^2\delta^2}. \quad (4)$$

The probability of false flagging now becomes:

$$\text{False flagging} = 2 * \text{F}(m_{het}(1 - d(\%)/100), m_{het}, s_{het}), \quad (5)$$

with m_{het} and s_{het} given by Eqs. 2 and 3, respectively. The larger δ becomes, the larger the false flagging rate and hence, the larger the effect of peer group heterogeneity on laboratory evaluation. When x_1, p and s_1 are known, Eq. 5 gives the relation between the degree of heterogeneity and the increase of false flagging. Equation 5 also enables the calculation of the maximal difference between subgroups δ when x_1, p and s_1 are known, and a limit on the increase of the false flagging rate has been proposed.

The logic was applied to the Croatian EQA scheme for general biochemistry. Subgroups were defined and heterogeneity was assessed by combining every two sets of subgroups that belonged to the same peer group. For every combination of subgroups, x_1 and s_1 were calculated as the mean and standard deviation of the largest subgroup and p was calculated as the proportion of the results that belonged to the other subgroup after outliers identified by a Grubbs test for which the p-value was smaller than 5%, were omitted. Using an increased allowed false flagging rate of 5% points, the parameter δ was derived from Eq. 5 and the number of samples that is needed to give a significant multivariate analysis of variance (MANOVA) test with a power of 90% if the difference is equal to δ was calculated [7–9]. A MANOVA test was applied to assess differences between subgroups, in which the multivariate response consisted of the results reported by the same laboratories for multiple samples.

For every group where the MANOVA test indicated that subgroups should be split, the improvement in flagging was considered. It can be assumed that, if subgroup size remains relatively large ($n \geq 10$), a flagging based on subgroup means is more correct than a flagging based on peer group means. For this reason, every result that is flagged based on the peer group mean but not based on the subgroup mean, and every result that is not flagged based on the peer group mean but is flagged based on the subgroup mean is considered as an improved evaluation. Peer groups were split only if the percentage of improved evaluations was higher than 5%.

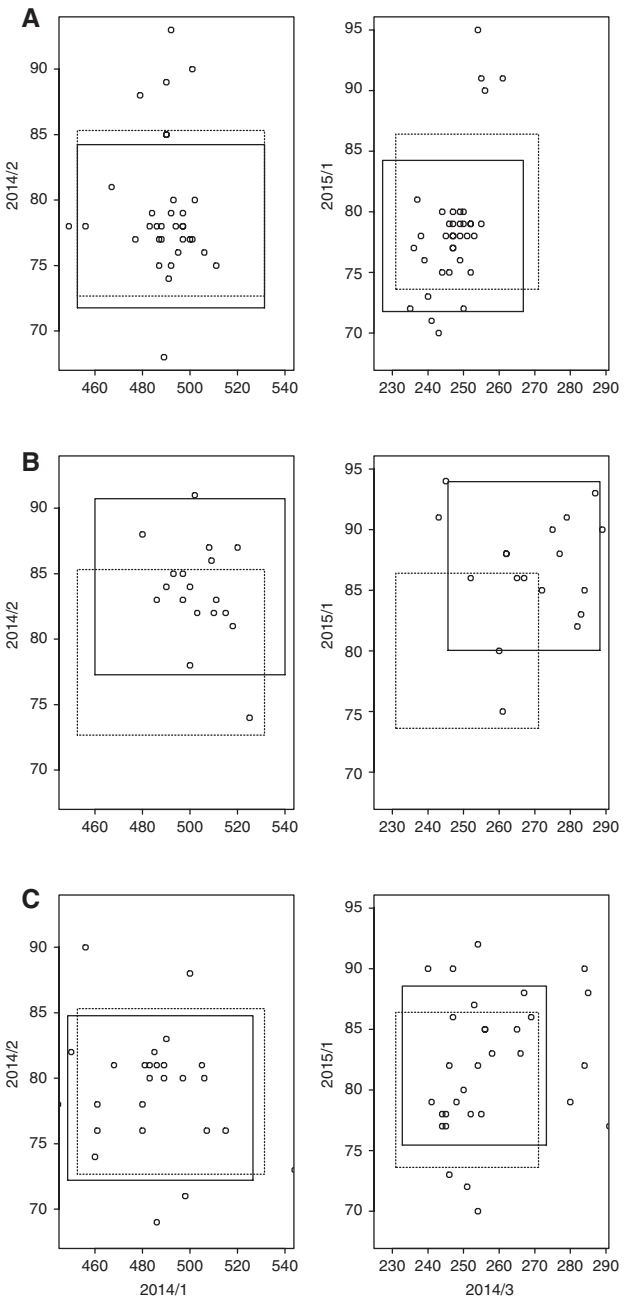


Figure 1: Graphical presentation of EQA results for creatinine, compensated Jaffe method traceable to NIST SRM 967 on (A) Beckman Coulter AU series (B) Roche Hitachi/Modular and (C) others. Left graph shows each time the results from round 2014/1 vs. 2014/2 and the right graph shows results from round 2014/3 vs. 2015/1. Rectangles within scatter plots represent ALP with respect to peer group (dotted rectangle) and subgroup (continuous rectangle) median. The position of rectangles to one another give indication that further splitting will improve evaluation of subgroups Roche Hitachi/Modular and Other.

Results

Twenty-nine method-based peer groups for 20 analytes were identified assuring each group has more than 10 participants and at least two instruments within same method used for analysis. For example, only two peer groups were created for sodium ('indirect ISE' and 'flame photometry') because 'direct ISE' peer group did not have enough participants to assure different instrument subgroups of at least 10 participants across four schemes. The groups were further divided into 75 subgroups according to instruments used for analysis. For each subgroup, data were visually explored by means of scatter plots where data of two

Laboratory ID	Analyzer manufacturer	Sample 1	Sample 2	Sample 3	Sample 4
1	Beckman Coulter AU series	490	85	247	80
7	Beckman Coulter AU series	497	78	250	72
8	Beckman Coulter AU series	501	77	249	78
(...)					
26	Roche Hitachi/Modular	503	86	289	90
53	Roche Hitachi/Modular	515	82	284	85
	Means per group				
	Beckman Coulter AU series	493	79.6	248	78.5
	Roche Hitachi/Modular	502	82.3	272	86.8

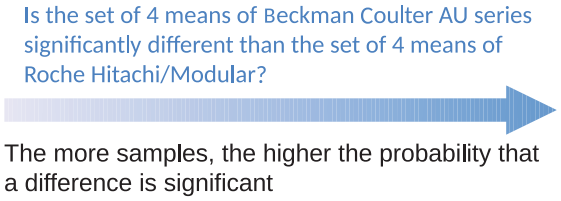


Figure 2: Workflow for MANOVA analysis applied to results for creatinine (compensated Jaffe method traceable to NIST SRM 967) in four rounds of Croatian EQA scheme.

successive EQA rounds were plotted against each other. Graphical presentation of EQA results with respect to ALP for each analyte revealed subgroups where further splitting might improve the laboratory evaluation (Figure 1).

For each of the subgroups, defined per analyte and method, mostly results from samples in two EQA rounds were selected based on the parameter δ from Eq. 5 to perform the MANOVA. For one group, three samples were needed, and for another group, four samples were needed to perform the MANOVA (Figure 2).

Based on statistically significant differences between pairs of different subgroups ($p < 0.05$), there was evidence for further splitting of 16 peer groups into subgroups. For example, reported results of MANOVA analysis for creatinine compensated Jaffe method traceable to NIST SRM 967 peer group are presented in Table 1. In each group for which the MANOVA test indicated that groups should be split, the amount of improved flagging was considered. The groups were actually split into subgroups if the improvement in laboratory evaluation exceeded 5% and clear improvement of correctly flagged results was observed for 6% of all results in groups needed to be split. For a particular method, this percentage was 27.4% (Roche Hitachi/Modular subgroup for creatinine). It should be noted though that the improvement was observed mainly in the subgroup 'Other' which is composed of different small groups of results. Table 2 summarizes the effect of peer group splitting applied to instrument subgroups where statistically significant differences between subgroups were found and clear improvement in flagging rate (above 5%) was observed in 50% of the cases (8/16).

Table 1: Results of MANOVA analysis for creatinine compensated Jaffe method traceable to NIST SRM 967 peer group.

Group comparison	p-Value for distinguishing groups
Beckman Coulter AU series – Other	0.0055
Beckman Coulter AU series – Roche Hitachi/Modular	< 0.0001
Other – Roche Hitachi/Modular	0.0001

Discussion and conclusions

Splitting peer groups is a trade-off between a large peer group with higher probability to be heterogeneous or small group with higher probability of unreliable summary statistics and evaluation, especially when split groups become small. Splitting peer groups according to manufacturer of equipment for a certain analyte is common in

Table 2: Effect of peer group splitting on correct flagging of laboratories' results.

Parameter	Method	Instrument	Number of data	Number of missed flags	Number of wrong flags	Percentage improved, %
Total bilirubin	DPD	Beckman Coulter AU	294	0	0	0.0
		Other	97	6	15	21.6
		Roche Hitachi/Modular	102	0	5	4.9
Creatinine	Compensated Jaffe, traceable to NIST SRM 967	Beckman Coulter AU	152	2	6	5.3
		Other	145	1	4	3.4
		Roche Hitachi/Modular	84	6	17	27.4
Creatinine	Jaffe	Beckman Coulter AU	129	1	1	1.6
		Other	139	5	5	7.2
Total protein	Biuret	Beckman Coulter AU	222	2	2	1.8
		Other	112	0	0	0.0
		Roche Cobas Integra	44	0	0	0.0
		Roche Hitachi/Modular	73	2	2	5.5
Chloride	ISE Indirect	Beckman Coulter AU	191	0	0	0.0
		Other	107	8	11	17.8
Albumin	BCG	Beckman Coulter AU	123	5	5	8.1
		Other	80	6	8	17.5
Total			2094	44	81	6

EQA schemes, but such an approach in Croatian scheme yields too many small peer groups, especially since lot of groups still show great heterogeneity in combinations of different manufacturers' instruments and reagents used. Creating larger, method-based peer groups seemed like a good solution to have more reliable statistic evaluation of particular result, but the obvious numerical dominance of one group (Beckman Coulter) shifted the target value toward median of its' results. Such an evaluation revealed many incorrect flagging for non-Beckman Coulter instruments. It was clear that a statistical approach has to be applied to find significant differences in consensus target values among different subgroups of instruments, with a general objective to separate peer groups that consist of different instrument-linked subgroups as soon as the within-subgroup variation is smaller than the between subgroup variation.

An important note should be made about significant differences between p-values. Significant p-value might not have any clinical relevance when a lot of data are involved, and a non-significant p-value may not lead to the right conclusions if too few data are involved. Since historic data from an EQA round cannot be augmented or reduced, a test is needed that involves multiple samples and for which the necessary number of involved samples can be calculated. For this reason, sample size planning prior to applying MANOVA test is a major prerequisite for obtaining correct and informative statistical analysis. The planning depends on the expected difference between subgroup means, subgroup size and the ratio between subgroup variability and the evaluation limit. It can be expected that false flagging of laboratories increases when peer groups are not split and subgroup means are situated further from each other. For this reason, the necessary number of samples were calculated in such a way that a theoretical increase in false flagging rate of 5% or more should give a significant MANOVA p-value with at least 90% of the time. Using this logic, the results of individual laboratory depend on the evaluation criteria in use and may change when new recommended analytical performance specifications are set [10].

The improvement in correct evaluation of EQA results is particularly high for creatinine, compensated Jaffe method traceable to NIST SRM 967. It might seem that standardization and traceability to primary reference material would yield a better harmonization among different instruments on which method is applied, but non-specificity in terms of interferences and cross-reactants questions the possibility of Jaffe method for creatinine to be standardized [11]. The measurement procedure that is sensitive to interference potentially introduces commutability problems even in minimally processed control samples, and the magnitude of interference depends on the individual sample [12]. Indeed, commutability is not

only the property of EQA sample, but rather combination of material-method interaction. The same non-specificity issues can be applied to uncompensated Jaffe method in creatinine determinations.

It has been shown previously that total bilirubin methods lack harmonization, particularly in the low concentration range [13]. We showed that harmonization was not achieved even within one method for total bilirubin (DPD method) probably because of lack of detailed explanation of DPD method and thus numerous variations in reagent composition and method design among different manufacturers [14].

Albumin concentration differences among different manufacturers can be explained by the chemically and immunologically undefined measurand, whereas differences in chloride measurements come as unpleasant surprise given that measurand is defined and standardization can be achieved through unbroken chain of traceability.

It should be stressed that peer group evaluation is recommended only when commutability of control material has not been validated and cannot be assumed due to numerous processing steps needed to ensure homogenous and stable samples with multiple and clinically relevant analyte composition and concentrations. All efforts from EQA providers should be directed toward creating commutable control material in order to enable target value assignment by reference methods and commutable materials where available. Consequently, such schemes can be used to monitor standardization and give valuable information about the harmonization of measurement procedures. Although this approach is preferable, commutability cannot always be achieved. Whenever commutability is unknown, the model can be of particular value, certainly when higher-order reference materials or methods [15] are not available (for example peptide hormones, tumor markers).

Choosing peer groups with comparable results from different equipment providers must be done with outmost care considering measurand and method itself, number of instruments and statistical evaluation. A MANOVA-based test, with a prior sample size planning that is based on maximal allowable difference between groups helps identifying groups that should be split and confirms the homogeneity of existing peer groups. The test may be applied as well for a continuous monitoring of the peer group homogeneity and for a fast detection of possible upcoming peer group heterogeneity, for which splitting would become recommended. Depending on the scheme design, a time period should be chosen by EQA providers for new evaluation of homogeneity of peer groups to allow any new potential subgroups to be created and/or to prevent splitting groups if differences among instruments enable correct evaluation within larger peer group. It should also be performed every time a new control material with different preparation protocol is introduced, given that the matrix related bias on individual measurement procedure might also be different. Whenever new group heterogeneity appears, caused by deviating results from individual instrument-based peer subgroup, EQA organizers should check if the observed difference is due to matrix effect or method bias.

Identifying heterogeneity within the peer group and splitting the groups accordingly enables verification that the individual laboratory is performing in accordance to the manufacturer's specifications and to other laboratories using the same technology. Observed differences among instrument-based subgroups can further be assessed by EQA providers for potential non-commutability of control samples, lack of harmonization or unsatisfactory accuracy of measurement procedure used if target values according to reference method can be assured. Until the origin of such differences is identified, observed heterogeneity within Croatian EQA biochemistry scheme lead us to decide that splitting specific peer groups led to a better laboratory evaluation.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: None declared.
Employment or leadership: None declared.
Honorarium: None declared.
Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. Dallas JG. Analytical performance specifications for EQA schemes – need for harmonisation. Clin Chem Lab Med 2015;53:919–24.
2. Miller WG, Jones GR, Horowitz GL, Weykamp C. Proficiency testing/external quality assessment: current challenges and future directions. Clin Chem 2011;57:1670–80.
3. Plebani M. External quality assessment programs: past, present and future. Jugosl Med Biohemija 2005;24:201–6.
4. Miller WG. Specimen materials, target values and commutability for external quality assessment (proficiency testing) schemes. Clin Chim Acta 2003;327:25–37.
5. Stepman HC, Tiikkainen U, Stockl D, Vesper HW, Edwards SH, Laitinen H, et al. Measurements for 8 common analytes in native sera identify inadequate standardization among 6 routine laboratory assays. Clin Chem 2014;60:855–63.

6. Flegar-Meštrić Z, Nazor A, Parag G, Sikirica M, Perkov S, Juretić D. Analytical goal-setting in external quality assessment for medical biochemistry laboratories in the Republic of Croatia (Ciljevi analitičke kvalitete u vanjskoj procjeni kvalitete rada medicinsko-biokemijskih laboratorija u Republici Hrvatskoj). Biochem Medica 2005;15:15–25.
7. O'Brien RG, Muller KE. Unified power analysis for t-tests through multivariate hypotheses. In: Edwards L, editor. Applied analysis of variance in behavioral science. CRC Press, 1993:297–344.
8. Olson CL. Practical considerations in choosing a MANOVA test statistic: a rejoinder to Stevens. Psychol Bull 1979;86:1350–2.
9. Olson CL. On choosing a test statistic in multivariate analysis of variance. Psychol Bull 1976;83:579.
10. Sandberg S, Fraser CG, Horvath AR, Jansen R, Jones G, Oosterhuis W, et al. Defining analytical performance specifications: consensus statement from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine. Clin Chem Lab Med 2015;53:833–5.
11. Koumantakis G. Traceability of measurement results. Clin Biochem Rev 2008;29(Suppl 1):S61–6.
12. Ceriotti F. The role of external quality assessment schemes in monitoring and improving the standardization process. Clin Chim Acta 2014;432:77–81.
13. Koerbin G, Tate JR, Ryan J, Jones GR, Sikaris KA, Kanowski D, et al. Bias assessment of general chemistry analytes using commutable samples. Clin Biochem Rev 2014;35:203.
14. Schlebusch H, Axer K, Schneider C, Liappis N, Röhle G. Comparison of five routine methods with the candidate reference method for the determination of bilirubin in neonatal serum. J Clin Chem Clin Biochem Z Für Klin Chem Klin Biochem 1990;28:203–10.
15. International Bureau of Weights and Measures. JCTLM database: laboratory medicine and in vitro diagnostics: database of higher order reference materials and reference measurement procedures. Available from: http://www.bipm.org/jctlm/. Accessed: 1 Jun 2016.

analytical chemistry

Article

Cite This: *Anal. Chem.* 2019, 91, 1847−1854

pubs.acs.org/ac

Alternative Sample-Homogeneity Test for Quantitative and Qualitative Proficiency Testing Schemes

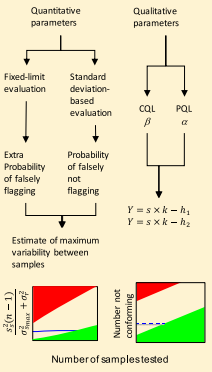
Wim Coucke,*[,†] Jelena Vlašić Tanasković,[‡] Lobna Bouacida,[†] Sylvia Broeders,[†] Bernard China,[†] Marianne Demarteau,[†] Vanessa Ghislain,[†] Yolande Lenga,[†] Marjan Van Blerk,[§] Nathalie Vandevelde,[†] Hannelien Verbeke,[†] Sandra Wathlet,[†] and Mohamed Rida Soumali[†]

[†]Sciensano, J. Wytsmanstraat 14, 1050 Brussels, Belgium

[‡]Department of Laboratory Diagnostics, General Hospital Pula, Zagrebačka 30, 52100 Pula, Croatia

[§]AZ Jan Portaals, Gendarmeriestraat 65, 1800 Vilvoorde, Belgium

ABSTRACT: Proficiency Testing (PT) External Quality Assessment (EQA) schemes are designed to ascertain the ability of individual laboratories to perform satisfactorily with respect to their peer laboratories or to limits imposed by external sources. Observed deviation of a laboratory result for a PT sample must be entirely attributed to the laboratory and not to the PT provider. To minimize the probability that deviations could be attributed to the PT provider, sample homogeneity should be assured. It is generally required that for quantitative parameters, the standard deviation among PT units should be calculated on the basis of duplicate measurements of at least 10 samples chosen at random, and the standard deviation among PT units should not exceed 0.3 times the standard deviation used to evaluate laboratories. Because this approach has important drawbacks, an alternative procedure is proposed by applying the theory of acceptance sampling to the assessment of sample heterogeneity for both quantitative and qualitative data and deriving acceptance limits on the basis of minimizing the probability of falsely evaluating laboratories. For obtaining acceptance limits for quantitative parameters, a distinction is made between laboratory evaluation using fixed limits on the one hand and laboratory evaluation using limits that are based on the variability of the reported results on the other hand. Sequential tests are proposed to evaluate sample heterogeneity by means of a comparison with the χ^2 distribution. For qualitative parameters, acceptance-sampling plans are proposed that are based on minimizing the joint probability of rejecting batches that have a satisfactory amount of defective samples and accepting batches unnecessarily. The approach for quantitative parameters is applied on samples for a PT scheme of ethanol quantification and for qualitative parameters on the presence of monoblasts in a blood smear. It was found that five samples could already be enough to prove that the batch was homogeneous for quantitative parameters, although more than 20 samples were needed to prove homogeneity for qualitative parameters. This study describes a direct relation among the objective of an PT round, the criteria for evaluating the results, and the sample heterogeneity. When samples are effectively homogeneous, less measurements are needed than current practices require. A drawback of the proposed approach is that the number of samples to be tested is not known beforehand, and good knowledge of the analytical variability is crucial. The formulas to be applied are relatively simple. Despite the drawbacks, the proposed approach is generally applicable for both quantitative and qualitative data.



Quality control in laboratory analysis includes different aspects. Apart from internal controls, external controls such as proficiency testing (PT) and external quality assessment (EQA) are designed to ascertain the ability of individual laboratories to perform satisfactorily with respect to their peer laboratories or to limits imposed by external sources. PT and EQA are often used interchangeably. The former is more often used in North America and is often related to regulatory or legal attributes. The latter is more often used within European areas, is often seen as a broader activity, of which laboratory evaluation makes up only a part and is usually regarded as educational. In this manuscript, the term PT will be used to describe both PT and EQA.

The general organization of a PT scheme is to distribute samples with the same content to the participating laboratories from the PT provider. Participating laboratories are asked to analyze the sample and send back the results to the PT provider. By comparing the results of each laboratory with the target value, laboratories are evaluated. Because the results of the PT are often part of an accreditation process, it is very important that the PT is managed correctly. Among other concerns, the observed deviation of a laboratory result from a PT sample must be entirely attributed to the laboratory and not to the PT provider. To minimize the probability that deviations could be attributed to the PT provider, PT samples and their content should be identical. International standards for PT schemes[1,2] require that for quantitative parameters, the standard deviation among samples should be calculated on the basis of duplicate measurements of at least 10 samples chosen

Received: July 24, 2018
Accepted: January 2, 2019
Published: January 4, 2019

DOI: 10.1021/acs.analchem.8b03313
Anal. Chem. 2019, 91, 1847−1854

at random, and formulas are provided to estimate the analytical repeatability, σ_r, and the standard deviation among PT samples, σ_s. In short, the estimate, s_S, of the variability between PT samples, σ_s, should be smaller than $0.3\sigma_p$, the PT standard deviation for a homogeneous sample. The notion of inference around the standard deviation among PT units and a proposal to apply an *F*-test instead of the simple $<0.3\sigma_p$ criterion have been introduced[3] and been adopted by the IUPAC international harmonized protocol for proficiency testing[2] and ISO 13528.[1] For qualitative parameters, the requirements are vague. ISO 13528,[1] for example, requires that an appropriate number of samples should have the right outcome, without specifying what an appropriate number of samples might be.

Although widely applied, the approach for quantitative parameters has some important drawbacks. The factor of 0.3 to be multiplied by σ_p was chosen to limit the contribution of the standard deviation among samples to less than 10% of the total variation of the reported results. As a consequence, *Z*-scores would be inflated to 5% or less.[3] It should be noted that this rule applies solely to cases in which a predefined standard deviation is used to calculate *Z*-scores, not the standard deviation of the reported PT results. In addition, because the calculation of s_S involves the square root of a difference that is more likely to be negative when the ratio of the repeatability, σ_r, over the sample variability, σ_s, increases, the calculation of s_S is not always possible, and the only solution is to consider σ_s equal to 0. Lastly, the inference test about σ_s rejects the sample only when there is clear proof of heterogeneity and may accept samples for which there is no proof that they are satisfactorily homogeneous. These arguments demonstrate that valid alternative approaches should be envisaged.

This study introduces the theory of acceptance sampling to the assessment of sample heterogeneity for both quantitative and qualitative data using data from sample preparations for ethanol in blood and lymphocyte subset counting, respectively. The theory of acceptance sampling, which originated more than a century ago,[4] was, in its early years, based on qualitative testing: a limited number of samples were taken from a batch, and the batch was rejected when the number of nonconforming units was above a predefined limit and accepted otherwise. Later, models for application to quantitative parameters were developed as well.[4] In order to determine the limits within which acceptance sampling for quantitative parameters should operate, the effects of sample heterogeneity on PT standard deviations and laboratory evaluations have to be outlined first.

CONTROLLING SAMPLE HETEROGENEITY FOR QUANTITATIVE RESULTS

Effect of Sample Heterogeneity on the Variability of Reported PT Results. Sample heterogeneity has a direct effect on the variability of the reported results. Without aiming at quantifying each of the individual components, it can be stated that the variance of the reported results, σ_s^2, consists of the sum of the variance among samples, σ_s^2, and the interlaboratory reproducibility, σ_R^2:

$$\sigma_p^2 = \sigma_s^2 + \sigma_R^2 \quad (1)$$

This equation assumes that all laboratories show the same analytical variability, a realistic assumption when the same analytical methodology is used across laboratories. It also assumes that variances can be summed up without taking into account covariances, a technique that is quite common in the field.[5] In fact, sample heterogeneity contributes to the total PT standard deviation in the same way as the interlaboratory reproducibility, σ_R: it will inflate the estimation of the PT standard deviation. If a correct estimation of the PT standard deviation is of interest, a limit on the inflation of the PT standard deviation by sample heterogeneity can be used to calculate a limit for the sample heterogeneity. For example, when inflation of the PT standard deviation induced by the sample heterogeneity should not exceed a certain proportion, *a* (e.g., 0.1, which equals 10%), by rearranging eq 1, we have

$$\sigma_{s_{max}} = \sqrt{a(a+2)\sigma_R^2} \quad (2)$$

with $\sigma_{s_{max}}$ being the maximum allowed value of σ_s. The term σ_R can also be considered as the standard deviation of the reported PT results for homogeneous samples. An estimate could be obtained from past PT rounds using homogeneous samples.

Effect of Sample Heterogeneity on Laboratory Evaluation. An increased PT standard deviation that is due to sample heterogeneity has diverse effects on laboratory evaluation, depending on how laboratories are evaluated. Two distinct types of laboratory evaluation exist: (a) using fixed limits, for example, a maximum allowable relative deviation from the target or consensus value, and (b) using limits that depend on the variability of the reported results.

Effect of Sample Heterogeneity on Laboratory Evaluation Based on Fixed Limits. When fixed limits are applied, a laboratory is flagged for poor performance when its reported result falls outside the interval $[L, U] = [x_a \pm dx_a]$ or outside the interval $[x_a \pm kx\sigma_p]$ when the PT standard deviation, σ_p, is fixed and known beforehand.[6] Parameters x_a, *d*, and *k* stand, respectively, for the assigned value, the maximum allowable relative deviation from x_a, and a percentile score (chosen by the PT organizer, usually 2 or 3) that reflects the number of standard deviations that a reported result is allowed to deviate from the assigned value. The effect of sample heterogeneity on a fixed-limits evaluation is displayed in Figure 1.

When the sample is homogeneous, σ_s is equal to 0, and hence the standard deviation of the reported results, denoted by $\sigma_{p|hom}$, is given by σ_R.

Even when laboratories perform well, there is a small probability that the value they reported is situated outside of the interval $[L, U]$. Let us call this probability the probability of falsely flagging. The term *false* refers to situations in which well-performing laboratories are flagged despite good performance. The probability of falsely flagging for a homogeneous sample ($p_{FF|hom}$) is given by

$$p_{FF|hom} = 2F(L, x_a, \sigma_{p|hom}) = 2 - 2F(U, x_a, \sigma_{p|hom}) \quad (3)$$

where *F* stands for the cumulative probability function of normally distributed values with x_a as the mean and a standard deviation of $\sigma_{p|hom}$.

When the sample is heterogeneous, σ_s is not equal to 0, and the standard deviation of the reported results can now be described by $\sigma_{p|het}$:

$$\sigma_{p|het} = \sqrt{\sigma_{p|hom}^2 + \sigma_s^2} \quad (4)$$

In this case, the probability of falsely flagging ($p_{FF|het}$) under heterogeneity is given by

DOI: 10.1021/acs.analchem.8b03313
Anal. Chem. 2019, 91, 1847−1854

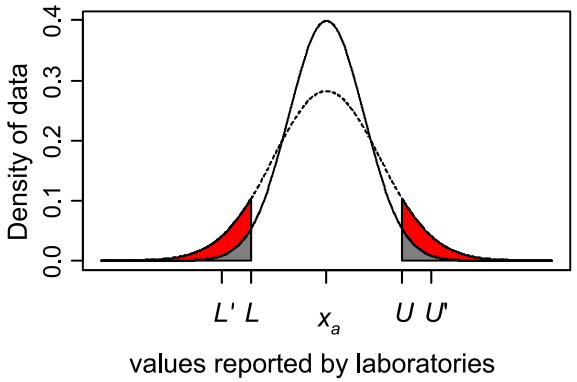


Figure 1. Effect of sample heterogeneity on laboratory evaluation using fixed limits. A theoretical distribution of reported data is shown for a homogeneous (solid line) and a heterogeneous (dashed line) sample, presuming a Gaussian distribution, around the assigned value (x_a), of values reported by well-performing laboratories. L and U are the limits beyond which laboratories are flagged for bad performance. The area under the curve beyond these values is called the probability of false flagging. It is indicated by the gray zones when the sample is homogeneous. The red zones show the increase in the probability of false flagging under sample heterogeneity. U' and L' are the limits for a heterogeneous sample with the same probability of falsely flagging as L and U in the case of a homogeneous sample.

$$p_{\mathrm{FF|het}} = 2F(L, x_a, \sigma_{\mathrm{p|het}}) = 2 - 2F(U, x_a, \sigma_{\mathrm{p|het}}) \quad (5)$$

The increase in the probability of falsely flagging due to heterogeneity ($\Delta p_{\mathrm{FF|het}}$) is given by

$$\Delta p_{\mathrm{FF}} = p_{\mathrm{FF|het}} - p_{\mathrm{FF|hom}} \quad (6)$$

Controlling the Effect on Laboratory Evaluation with Fixed Limits. The maximum allowable sample heterogeneity, $\sigma_{s_{max}}$, can be calculated according to two methods. The first method consists of putting limits on Δp_{FF} by defining an allowed probability of falsely flagging. The maximum allowed sample heterogeneity, $\sigma_{s_{max}}$, can be derived by an iterative process, in which $\sigma_{s_{max}}$ is estimated such that Δp_{FF} remains limited. The process consists of deriving Δp_{FF} from eqs 3−5 by using a starting value for $\sigma_{s_{max}}$. When the obtained Δp_{FF} is above the limit, the whole process is repeated for a small $\sigma_{s_{max}}$, and when the obtained Δp_{FF} is below the limit, the whole process is repeated for a larger $\sigma_{s_{max}}$. Algorithms exist to calculate the optimum values of $\sigma_{s_{max}}$ after each iteration step.[7]

The second approach of putting limits on the flawed laboratory evaluation consists of considering the actual deviation that gives the same probability of falsely flagging as under sample homogeneity, which can be denoted by $d(\%)'$. It is found by calculating $d(\%)'$ such that

$$F(L', x_a, \sigma_{\mathrm{p|het}}) = F(L, x_a, \sigma_{\mathrm{p|hom}}) \quad (7)$$

where L' is the acceptance limit under sample heterogeneity that leads to the same probability of falsely flagging as L does under sample homogeneity. L could be replaced by U, and L' could be replaced by U' in this equation (see Figure 1). Here, an iterative calculation has to be used to obtain $d(\%)'$ after choosing an initial estimate of $\sigma_{s_{max}}$.

Effect on Laboratory Evaluation Based on Standard-Deviation-Dependent Limits. When laboratories are evaluated on the basis of the number of standard deviations that their reported value deviates from the assigned value, and the standard deviation is calculated from the reported results, there is no fixed interval of acceptability, and limits for flagging laboratories extend when the standard deviation of the reported results increases. This means that compared with an evaluation on fixed limits, sample heterogeneity has another effect on the laboratory evaluation: a portion of the results are not flagged, whereas they would have been flagged if the sample had been homogeneous. As a consequence, if the sample is heterogeneous, the probability of falsely flagging, on the one hand, remains unchanged, and the probability of falsely not flagging, on the other hand, appears. The relation between the probability of falsely not flagging (p_{FNF}) and the PT standard deviation, calculated on the basis of the reported results, is illustrated in Figure 2.

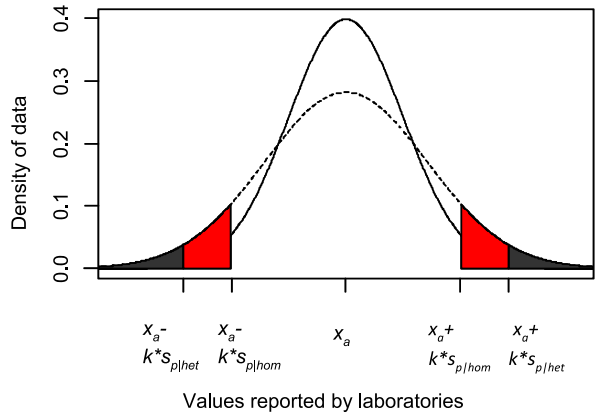


Figure 2. Effect of sample heterogeneity on laboratory evaluation using standard-deviation-dependent limits. The solid line shows the distribution of reported PT results for a homogeneous sample; the dashed line shows the distribution for a heterogeneous sample. The black zones show the probability of falsely flagging under sample homogeneity; the red zones show the probability of falsely not flagging under sample heterogeneity. The factor k stands for the number of standard deviations that a reported result is allowed to deviate from the assigned value; k is usually 2 or 3.

Note that that the probability of falsely not flagging results appears only in the case of heterogeneity. For this reason, contrary to the case of flagging laboratories with fixed limits, no Δ symbol is required, and the subscript *het* does not need to be used to describe the probability of falsely not flagging. If the assigned value is represented by x_a, the expected standard deviation of the reported results from a homogeneous sample is represented by $\sigma_{\mathrm{p|hom}}$, and that of a heterogeneous sample is represented by $\sigma_{\mathrm{p|het}}$, the following holds:

$$p_{\mathrm{FNF}} = 2F(x_a - k\sigma_{\mathrm{p|hom}}, x_a, \sigma_{\mathrm{p|het}}) - 2F(x_a - k\sigma_{\mathrm{p|het}}, x_a, \sigma_{\mathrm{p|het}}) \quad (8)$$

The proportion of well-performing values that are flagged when the sample is heterogeneous is given by the last part of eq 8. This proportion depends solely on k and is equal to 0.0455 when $k = 2$ and to 0.0027 when $k = 3$.

Controlling the Effect on Laboratory Evaluation with Standard-Deviation-Dependent Limits. The probability of falsely not flagging is chosen beforehand, and $\sigma_{\mathrm{p|hom}}$ can be estimated from previous PT results. When k is chosen (usually

DOI: 10.1021/acs.analchem.8b03313
Anal. Chem. 2019, 91, 1847−1854

2 or 3), the PT standard deviation that relates to the maximum probability of falsely not flagging can be obtained from eq 8 by

$$F(x_a - k\sigma_{plhom}, x_a, s_{plhet}) = 0.5p_{FNF} + F(x_a - k\sigma_{plhet}, x_a, \sigma_{plhet}) \quad (9)$$

By rewriting the left part of eq 9 as $F\left(\frac{x_a - k\sigma_{plhom} - x_a}{\sigma_{plhet}}, 0, 1\right)$, we obtain

$$F\left(-k\frac{\sigma_{plhom}}{\sigma_{plhet}}, 0, 1\right) = 0.5p_{FNF} + F(x_a - k\sigma_{plhet}, x_a, \sigma) \quad (10)$$

Once a limit for p_{FNF} is chosen, the right part of eq 9 is independent from the variability of the reported results. Let us define Q as the value for which the area to the left of Q under a standard normal curve is equal to $0.5p_{FNF} + F(x_a - k\sigma_{plhet}, x_a, \sigma_{plhet})$.

We then have

$$k\frac{\sigma_{plhom}}{\sigma_{plhet}} = -Q \quad \text{or} \quad \sigma_{plhet} = -\sigma_{plhom}\frac{k}{Q} \quad (11)$$

After that, the allowed sample heterogeneity can be obtained by means of eq 4:

$$\sigma_{s_{max}} = \sqrt{\sigma_{plhet}^2 - \sigma_{plhom}^2} = \sigma_{plhom} \times \sqrt{\frac{k^2}{Q^2} - 1} \quad (12)$$

with $\sigma_{s_{max}}$ being the maximum allowable value of σ_s.

Another way to calculate the maximum allowed sample heterogeneity is by considering the effective Z-limit, which is the limit that gives the same probability of falsely flagging when the sample is heterogeneous as the original Z-limit yields when the sample is homogeneous. It is given by k_e, the effective value of k, and is calculated by

$$k_e = k\frac{\sigma_{plhet}}{\sigma_{plhom}} \quad (13)$$

After proposing a value for k_e, eq 4 can be used to calculate $\sigma_{s_{max}}$.

Estimating Expected PT Standard Deviation under Homogeneity. Independently from how laboratories are evaluated, the expected PT standard deviation when the sample is homogeneous, σ_{plhom}, needs to be estimated, and PT organizers can rely on PT results reported in the past. Eventually, a supposed estimated sample heterogeneity could be subtracted from the calculated PT standard deviations from the past. Because the PT standard deviation of the reported results varies from sample to sample, an interpolation between standard deviations for different concentrations is recommended, for example, by using the characteristic function,[8,9] a function that draws the relation between the target value and standard deviation.

Evaluating Sample Heterogeneity to Keep Wrong Evaluations under Control. With the introduction of sequential tests,[10] it became clear that acceptance sampling by analyzing a predefined number of samples could be performed in a more efficient way, with, on average, fewer samples. Sequential-sampling plans were introduced, which consist of sampling unit by unit, calculating an evaluation statistic, and comparing this statistic with upper and lower limits; the sample is rejected when the evaluation statistic exceeds the upper limit and accepted when the statistic is below the lower limit. As long as the evaluation statistic is between the two limits, testing of an extra unit should be performed, with a new evaluation statistic calculated and evaluated with respect to the two limits.

Sequential tests can be performed to evaluate sample heterogeneity by means of a comparison with the χ^2 distribution. It is built on the basic idea that when the sample heterogeneity is exactly equal to the maximum allowed sample heterogeneity, the following equation holds:

$$\frac{s_s^2(n-1)}{\sigma_{s_{max}}^2 + \sigma_r^2} \sim \chi_{n-1}^2 \quad (14)$$

where σ_r is the analytical variability of the method that is used for evaluating the sample homogeneity, $\sigma_{s_{max}}$ is the maximum allowed sample heterogeneity (both expressed as standard deviations), n is the number of vials taken so far, and s_s is the standard deviation of the samples that are evaluated.

Note that all variability estimates in eq 14 are written as variances (i.e., squares of the standard deviations). When the actual standard deviation between the samples is smaller than $\sigma_{s_{max}}$, the ratio will be smaller than expected under a χ^2 distribution, and when the actual standard deviation is larger, the ratio will be larger than expected under a χ^2 distribution. The evaluation of sample heterogeneity is explained in Figure 3.

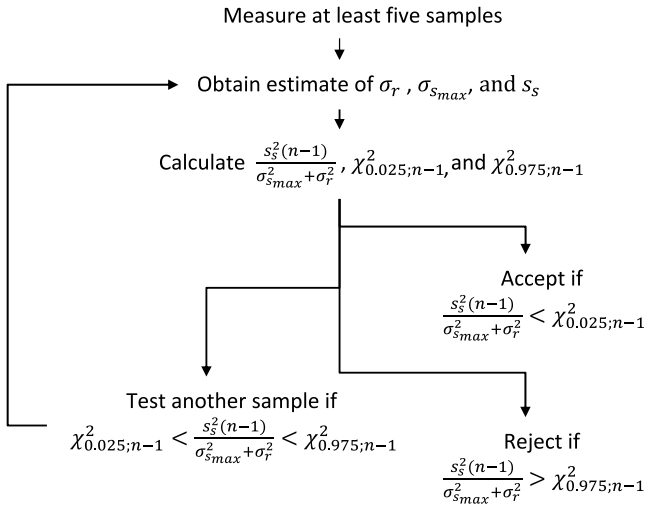


Figure 3. Flowchart for evaluating sample heterogeneity for quantitative parameters.

Initially, five vials are measured, and their mean is used to derive σ_r from measurements obtained during method validation; σ_r is then used to calculate the ratio $\frac{s_s^2(n-1)}{\sigma_{s_{max}}^2 + \sigma_r^2}$, and this ratio is compared with a χ^2 distribution with $n - 1$ degrees of freedom. Testng is stoppend and the batch is accepted when the ratio $\frac{s_s^2(n-1)}{\sigma_{s_{max}}^2 + \sigma_r^2}$ is located in the left 2.5% tail of the χ^2 distribution and the batch is rejected when it is located in the right 2.5% tail of the χ^2 distribution. Testing continues when the ratio $\frac{s_s^2(n-1)}{\sigma_{s_{max}}^2 + \sigma_r^2}$ is located in the middle 95% part of the

DOI: 10.1021/acs.analchem.8b03313
Anal. Chem. 2019, 91, 1847−1854

distribution, and after each analyzed sample, the ratio $\frac{s_s^2(n-1)}{\sigma_{s_{max}}^2 + \sigma_r^2}$ and the χ^2 distribution are updated and evaluated.

Controlling Sample Heterogeneity for Qualitative Results. A test for sample heterogeneity for qualitative results can be constructed using the operator-characteristic curve. It is illustrated in Figure 4.

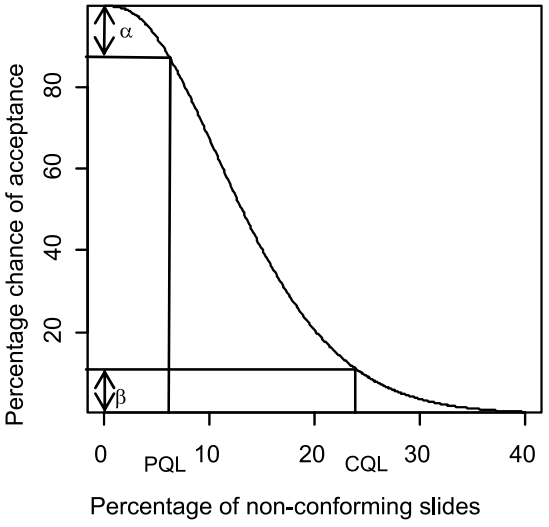


Figure 4. Example of an operator-characteristic curve. The producer's quality level (PQL) is the percentage of nonconforming units below which a producer does not want to have batches rejected, with probability $1 - \alpha$, and the consumer's quality level (CQL) is the percentage of nonconforming units above which a consumer does not want to have batches accepted, with probability $1 - \beta$.

The curve illustrates the decisions that can be taken when a fixed-sized sample of units is taken from a large batch, and each unit in the sample is tested for conformity.

The proportion of nonconforming units in the sample reflects the true proportion of nonconforming units in the whole batch and can be used to decide whether the batch should be rejected or accepted: the batch is rejected when the number of nonconforming units is too large, and it is accepted otherwise. Because the proportion of nonconforming units in the sample is only a reflection and hence only gives an approximate value of the proportion in the whole batch, batches may be accepted or rejected wrongly when the approximation is not good enough.

Two types of risks play a role in this decision process. When the sample contains a number of nonconforming units that is high with respect to the total nonconforming units in the batch, the batch may be wrongly rejected. When the sample contains a number of nonconforming units that is low with respect to the total nonconforming units, the batch may be wrongly accepted.

From the point of view of laboratory evaluation, it is disadvantageous to accept batches with too many nonconforming units. The quality level, expressed as the proportion of nonconforming units in the batch, beyond which a correct laboratory evaluation is jeopardized is called the consumer's quality level (CQL), and the probability of accepting a batch with a worse quality level is called β. From the point of view of costs involved in sample preparation, it is disadvantageous to reject batches unnecessarily. The quality level below which the sample provider does not want to have batches rejected is called the producer's quality level (PQL), and the chance of rejecting a batch with a proportion of nonconforming units below PQL is α. Evidently, PQL is always lower than CQL.

Similar to the evaluation of the sample heterogeneity for quantitative parameters, sequential testing can be applied for qualitative parameters as well. On the basis of predefined PQL, CQL, α, and β, a rejection–acceptation graph can be drawn as shown in Figure 5 using the following parameters:[4,9]

$$h_1 = \frac{b}{G},\ h_2 = \frac{a}{G},\ s = \frac{g_2}{G}$$

where $a = \log\left(\frac{1-\beta}{\alpha}\right)$, $b = \log\left(\frac{1-\alpha}{\beta}\right)$, $g_1 = \log\left(\frac{\mathrm{CQL}}{\mathrm{PQL}}\right)$, $g_2 = \log\left(\frac{1-\mathrm{PQL}}{1-\mathrm{CQL}}\right)$, and $G = g_1 + g_2$.

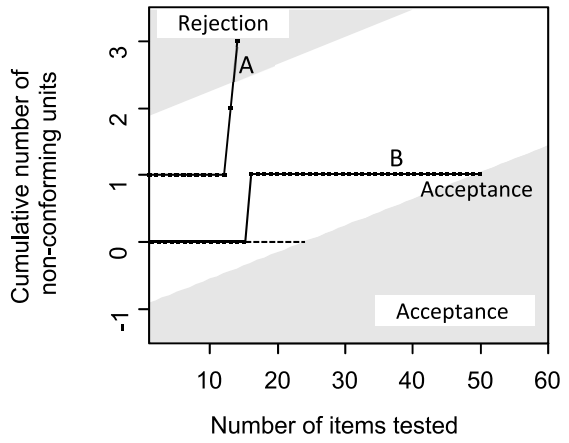


Figure 5. Rejection and acceptance zones for a PQL of 1%, a CQL of 10%, an α of 1%, and a β of 10%. Two lines are drawn to illustrate two distinct testing processes, A and B. Testing process A starts with a nonconforming unit, after which testing continues. Eleven conforming units follow and make the line stretch horizontally. The 13th and 14th units, however, are nonconforming, which makes the line to reach the rejection zone, after which the batch is rejected. Process B starts with 15 consecutive conforming units. The 16th unit is nonconforming, which makes the line to shift up one unit. All subsequent units are conforming, which makes it reasonable to accept the batch after testing 51 units.

The parameters h_1, h_2, and s are used to calculate two lines:

line of acceptance

$$Y = s \times k - h_1$$

line of rejection

$$Y = s \times k - h_2$$

where Y stands for the cumulative number of nonconforming vials, and k stands for the number of vials tested so far.

Randomly chosen units are consecutively tested, and the cumulative number of nonconforming units is plotted on the graph. The sample is accepted when the cumulative number of nonconforming units (Y) is smaller than the line of acceptance and rejected when Y is larger than the line of rejection.

The graph is equipped with rejection and acceptance zones, of which the positions and the spaces are determined by α, β, PQL, and CQL. Testing continues until the sample is rejected or accepted (see Figure 5).

DOI: 10.1021/acs.analchem.8b03313
Anal. Chem. 2019, 91, 1847−1854

Table 1. PT Standard Deviation, Maximum Sample Variability, Analytical Variability, and Maximum Number of Data Points Needed for the Five Samples Evaluated for Homogeneity for Serum Ethanol

sample concentration (g/L)	lowest σ_{plhom} (g/L)	highest σ_{plhom} (g/L)	analytical variability, σ_p (g/L)	relative-difference-linked maximum sample heterogeneity (g/L)	Z-score-linked maximum sample heterogeneity	maximum sample heterogeneity (g/L)
0.15	0.0070	0.0253	0.0041	0.008 14	0.006 21	0.006 21
0.45	0.0133	0.0289	0.0084	0.0155	0.0118	0.0118
0.75	0.0208	0.0367	0.0134	0.0328	0.0184	0.0184
1.5	0.0401	0.0734	0.0262	0.0655	0.035	0.0354
3	0.0762	0.1469	0.0521	0.1309	0.0672	0.0672

MATERIALS AND METHODS

The theory of acceptance sampling for quantitative parameters was applied on a PT scheme for serum ethanol. Five samples for a PT survey of ethanol were tested. The Belgian PT survey assesses laboratory performance in two ways: laboratories are flagged when their *Z*-scores (based on robust peer-group statistics) are beyond 3 or when the relative difference with respect to the consensus median is larger than 25% for samples with concentrations <0.4 g/L and larger than 15% for samples with concentrations ≥0.4 g/L.

Samples were made of fresh, non-alcohol-containing serum and were spiked to obtain concentrations of 0.15, 0.45, 0.75, 1.5, and 3 g/L ethanol, respectively. The analytical variability was determined on 20 different measurements for 7 different controls with concentrations ranging from 0.0617 to 3.950 g/L. PT standard deviations from the past were determined by means of the characteristic function on the basis of results of the last 33 samples, which were reported from 2013 to 2015. Only methods used by at least 10 participants were taken into account. As a result, five different analytical methods were taken into consideration. For the *Z*-scores, an effective limit of 4 was proposed and for the relative differences, an increase in the probability of falsely flagging of 2 percentage points was accepted.

The theory of acceptance sampling for qualitative parameters was applied to the evaluation of a blood smear for use in a PT for hematology. Randomly selected blood smears from a patient with acute monoblastic leukemia were evaluated for the percentage of monoblasts. PQL, CQL, α, and β, were set to 1, 20, 1, and 1% respectively. On every slide, 100 leucocytes were counted and a blood smear was considered as conforming when at least 60 monoblasts were identified.

RESULTS

In order to deal with the various values for the PT standard deviation, the highest and lowest PT standard deviations for various methods where calculated from the last 33 samples that were used in the Belgian PT scheme for serum ethanol, ranging from 0.14 to 2.99 g/L, using the characteristic function. The analytical variability for the specific sample concentrations was derived by interpolation on the characteristic function and is given in Table 1. It ranges from 0.0041 g/L for the lowest target concentration to 0.0521 g/L for the highest target concentration.

Table 1 shows the maximum allowed sample heterogeneity for relative differences, the maximum allowed heterogeneity for *Z*-scores, the final maximum allowed sample heterogeneity (the smallest of the two), and the analytical variability.

The evaluation of sample heterogeneity, as obtained by sequential testing, is listed in Table 2. Enough evidence for homogeneity was reached after five to seven measurements.

Table 2. Consecutive Measurements for Each Sample[a]

measurement	sample 1 (0.15 g/L)	sample 2 (0.45 g/L)	sample 3 (0.75 g/L)	sample 4 (1.5 g/L)	sample 5 (3 g/L)
1	0.13	0.45	0.73	1.51	2.99
2	0.13	0.46	0.74	1.48	2.99
3	0.13	0.46	0.73	1.47	3.05
4	0.13	0.45	0.73	1.49	2.97
5	0.13	0.45	0.73	1.47	3.05
6		0.47		1.52	3.05
7		0.46		1.49	2.97

[a]Continued until enough evidence was found that the sample could be considered sufficiently homogeneous.

For the qualitative testing, the sample could be accepted after evaluating 25 samples that all conformed. The acceptance and rejection regions, together with the line representing the 25 conforming results, are shown in Figure 6.

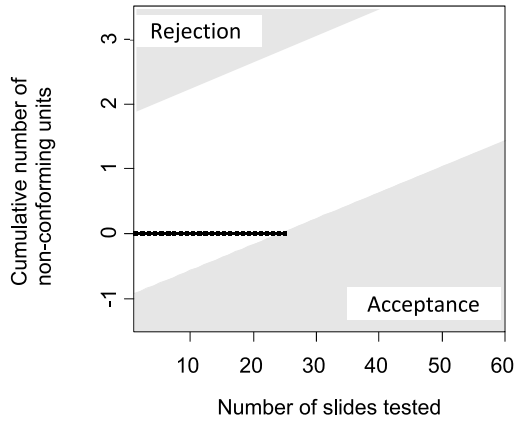


Figure 6. Acceptance and rejection zones for testing for monoblast counting. Because no nonconforming slides were found, the batch was accepted after the evaluation of 25 slides.

DISCUSSION

Because evaluation criteria may vary widely among PT organizers, deriving limits from sample heterogeneity that are more apt to the evaluation criteria than the classical criterion of $0.3\sigma_p$ is the first important issue raised in this study. The ISO 13528:2012 standard acknowledged that the classical criterion does not hold when the standard deviation is calculated on the basis of the reported results but offers very limited alternatives.

Whatever the limit for the standard deviation among samples, when a single-point estimator like the standard deviation among samples is compared with the limit, as the international standards currently request, there is a probability

DOI: 10.1021/acs.analchem.8b03313
Anal. Chem. 2019, 91, 1847−1854

that a batch will be wrongly accepted or rejected, and this probability increases when the actual sample heterogeneity approaches the acceptance limit. This study describes a direct relation among the objective of a PT round, the criteria for evaluating the results, and the sample heterogeneity.

Concerning quantitative data, evaluations based on fixed limits require that the chance that a good result would be falsely flagged (the probability of falsely flagging) has to be controlled. For evaluations that are based on limits using the standard deviation that is calculated using the reported data, the chance of not flagging a bad result (the probability of falsely not flagging) has to be controlled. For PT schemes that evaluate the laboratories by different methods, the sample heterogeneity requirements are not necessarily the same for the different evaluation methods, and the strictest requirement should be chosen. For the case of serum ethanol, the *Z*-score based on a standard deviation derived from the reported data required the lowest sample heterogeneity.

For qualitative data, a sampling plan is presented by building a protection against rejecting batches with very few nonconforming units and accepting batches with a lot of nonconforming units. The sampling plan is based on quantitated choices of "very few" and "a lot" and the degree of protection.

Concerning the relation between the allowed sample heterogeneity and PT standard deviation, this study shows that the limit of $0.3\sigma_p$ is not always appropriate for effectively reducing the chance of falsely flagging or not flagging laboratories. The example in this study has shown that the limits for variability between samples may even exceed the PT standard deviation for evaluations based on fixed limits. For evaluations based on limits based on the variability of reported results, we found that the standard deviation among samples may be up to $0.9\sigma_p$ without jeopardizing a correct laboratory evaluation in PT. Of course, these numbers depend not only on the tolerance with respect to a correct laboratory evaluation but also on the evaluation criteria and the expected variability of the reported results. Moreover, the theory is worked out for data that are normally distributed. The distribution of the data should be tested for every assessed parameter. We believe that one assessment based on reported PT results is sufficient for assessing the normal distribution.

The proposed methods for evaluating batches of PT samples for qualitative or quantitative data rely heavily on a random selection of vials, and it should be emphasized that in case the order of sample production is known, computer algorithms and not humans should define which vials are to be taken. In order to identify possible faults in the production process, a stratified sample may be considered, in which the complete production batch is divided into sub-batches of equal size, and a random selection per sub-batch should be considered. The algorithms however, do not take into account the order of sample preparation. In case heterogeneity could be linked to a trend in the sample preparation, for example when a certain sedimentation of cells would occur during the sample-production process, other techniques that are based on detecting a trend, using linear regression, could be worked out as well.

Current standards require at least 10 samples to be measured in duplicate without taking into account the difference between the actual sample heterogeneity and the acceptance limit. In fact, the number of measurements to be performed depends on the difference between the actual sample heterogeneity and its limit. The higher the difference, the lower the number of data that are needed. For the case study of serum ethanol described here, batches could already be accepted after testing five samples.

It should be borne in mind that all calculations performed here are made in the context of proving homogeneity for a sample preparation for which there is no evidence that it gives homogeneous samples and for which the preparation order is not known. Sample homogeneity for materials that can easily be mixed, like liquids, is much easier to accomplish, and hence, for a preparation method that has been demonstrated as yielding homogeneous samples, criteria can be less tight, requiring fewer samples to test for future preparations.

In contrast to the low number of samples for quantitative results and the small ratio between actual and allowable sample heterogeneity, the number of required samples for qualitative testing is higher. Even with moderately high values for PQL and CQL, more than 20 samples have to be tested before the sample can be accepted for homogeneity. If several samples are sent in one round and laboratories are only flagged if they report wrong results for more than one sample, higher PQL and CQL values could be used, and as a consequence, a lower number of samples have to be tested.

The requirements for sample heterogeneity also depend on the tolerance that a PT organizer has with respect to falsely evaluating laboratories by flagging good results or not flagging bad results. This tolerance may be used to rank PT schemes,[11] in which higher-ranked PT schemes that use commutable material, with target values set by reference methods, have lower tolerances for false flagging or nonflagging. In fact, this approach allows one to model the sensitivities for both the PT organizer and the PT participant. A PT organizer wants to avoid rejecting batches that are wrongly evaluated as too heterogeneous or having too much nonconforming units, whereas a PT participant wants to participate in PT schemes for which wrong laboratory evaluations are minimal and well-described. It suffices for PT providers to mention in their general description or in the reports which limits the samples have been tested against for homogeneity. Of course, all calculations made here and the results only apply to well-performing laboratories (i.e., laboratories that produce good PT results) and for which a flagging is rare and coincidental.

Two major drawbacks of this approach need to be reported. A first drawback is the a priori unknown number of vials to be tested. The sequential vial heterogeneity testing continues until enough evidence is found that the sample heterogeneity is smaller than the sample heterogeneity limit, after which the batch is accepted, or larger than the sample heterogeneity limit, after which the batch is rejected. The closer the actual sample heterogeneity and the limit are to each other, the more vials will be needed in order to collect enough evidence to accept or reject the batch. While applying the proposed methodology in testing homogeneity for PT samples, we experienced that up to 20 samples were needed to confirm sufficient homogeneity for quantitative parameters. Theoretically, this number could be higher and in the most extreme cases; when the actual sample heterogeneity is equal to the limit, an infinite number of vials are needed. Several solutions can be envisaged, for example by defining an actual and an upper limit of acceptability for the sample heterogeneity. Another solution consists of abandoning the idea of sequential testing. Although sequential testing has the advantage of exhibiting the highest efficiency, other

DOI: 10.1021/acs.analchem.8b03313
Anal. Chem. 2019, 91, 1847−1854

approaches may be easier to realize in the laboratory on automated analyzers, like double and single testing plans.[4]

The second drawback is the dependency on the knowledge of the analytical repeatability. When it is overestimated, heterogeneous samples could be wrongly accepted. Otherwise, homogeneous batches could be wrongly rejected when it is underestimated. Although testing analytical variability is an essential aspect of method validation, it is recommended that a measure of analytical variability that is based on at least 20 measurements is used and reassessed regularly, for example, by using the 50 last variability estimates of the controls. Another solution might consist of considering the variability of multiple measurements in one vial and the variability that is obtained just before performing the sample heterogeneity measurements.

Despite the drawbacks, the proposed approach is generally applicable, both for quantitative and qualitative data.

AUTHOR INFORMATION

Corresponding Author

*E-mail: wim.coucke@sciensano.be.

ORCID

Wim Coucke: 0000-0002-4290-4628

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors want to thank the Louvain Centre for Toxicology and Applied Pharmacology for performing the ethanol assays. This manuscript was written without any specific funding.

REFERENCES

(1) *Statistical Methods for Use in Proficiency Testing by Interlaboratory Comparisons*; ISO 13528:2015; International Organization for Standardization, 2015.

(2) Thompson, M.; Ellison, S. L.; Wood, R. *Pure Appl. Chem.* **2006**, *78* (1), 145−196.

(3) Fearn, T.; Thompson, M. *Analyst* **2001**, *126* (8), 1414−1417.

(4) Schilling, E. G. *Acceptance Sampling in Quality Control*; CRC Press, 1982.

(5) Iso, I.; Oiml, B. Guide to the Expression of Uncertainty in Measurement, 1995. *Bureau International des Poids et Mesures*. https://www.bipm.org/en/publications/guides/gum.html (accessed Nov 9, 2018).

(6) Coucke, W.; Soumali, M. R. *Biochemia medica* **2017**, *27* (1), 37−48.

(7) Nelder, J. A.; Mead, R. *Comput. J.* **1965**, *7* (4), 308−313.

(8) Coucke, W.; Charlier, C.; Lambert, W.; Martens, F.; Neels, H.; Tytgat, J.; Van de Walle, P.; Vanescote, A.; Wallemacq, P.; Wille, S.; et al. *Clin. Chem.* **2015**, *61* (7), 948−954.

(9) Thompson, M. *J. AOAC Int.* **2012**, *95* (6), 1803−1806.

(10) Wald, A. *Sequential Analysis*; Wiley: New York, 1947.

(11) Miller, W. G.; Jones, G. R.; Horowitz, G. L.; Weykamp, C. *Clin. Chem.* **2011**, *57* (12), 1670−1680.

DOI: 10.1021/acs.analchem.8b03313
Anal. Chem. 2019, 91, 1847−1854

Analytical verification of 12 most commonly used urine dipsticks in Croatia: comparability, repeatability and accuracy

Dora Vuljanić*[1], Ana Dojder[1], Valentina Špoljarić[1], Andrea Saračević[1], Lora Dukić[1], Jasna Leniček Krleža[2,5], Jelena Vlašić Tanasković[3,5], Ivana Maradin[4,5], Ana Grzunov[2,5], Željka Vogrinc[6], Ana-Maria Šimundić[1]

[1]Department of Medical Laboratory Diagnostics, University Hospital "Sveti Duh", Zagreb, Croatia
[2]Department of Laboratory Diagnostics, Children's Hospital Zagreb, Zagreb, Croatia
[3]Department of Laboratory Diagnostics, General Hospital Pula, Pula, Croatia
[4]Medical - biochemistry Laboratory: "Mirjana Plavetić and Ivana Maradin", Karlovac, Croatia
[5]Croatian Centre for Quality Assessment in Laboratory Medicine (CROQALM), Croatian Society of Medical Biochemistry and Laboratory Medicine, Zagreb, Croatia
[6]Department of Laboratory Diagnostics, University Hospital Centre Zagreb, Zagreb, Croatia

*Corresponding author: dora.vuljanic7@gmail.com

Abstract

Introduction: Variability among manufacturers of urine dipsticks, respective to their accuracy and measurement range, may lead to diagnostic errors and thus create a serious risk for the patient. Our aims were to determine the level of agreement between 12 most commonly used urine dipsticks in Croatia, examine their accuracy for glucose and total protein and to test their repeatability.

Materials and methods: A total of 75 urine samples were used to examine comparability and accuracy of 12 dipstick brands (Combur 10 TestM, ChoiceLine 10, Combur 10 TestUX, ComboStik 10M, ComboStik 11M, CombiScreen 11SYS, CombiScreen 10SL, Combina 13, Combina 11S, Combina 10M, UriGnost 11, Multistix 10SG). Agreement between each dipstick and the reference (Combur 10 TestM) was expressed as kappa coefficient (acceptable $\kappa \geq 0.80$). Accuracy for glucose and total protein was tested by comparison with quantitative measurements on analysers: AU400 (Beckman Coulter, USA), Cobas 6000 c501 (Roche Diagnostics, Germany) and Architect plus c4000 (Abbott, USA). Repeatability was assessed on 20 replicates (acceptable > 90%).

Results: Best agreement was achieved for glucose, total protein and nitrite (11/11, $k > 0.80$) and the lowest for bilirubin (5/5, $k < 0.60$). Sensitivities for total protein were 41-75% (AU400) and 56-92% (Cobas and Architect); while specificities were 41-75% (AU400, Cobas, Architect). Dipsticks' sensitivity and specificity for glucose were 68-98%. Most of the dipsticks showed unacceptable repeatability (6/12, < 90%) for one parameter, most prominently for pH (3/12, < 90%).

Conclusions: Most commonly used dipsticks in Croatia showed low level of agreement between each other. Moreover, their repeatability vary among manufacturers and their accuracy for glucose and proteins is poor.

Keywords: verification; urine dipsticks; comparability; accuracy; repeatability

Received: September 26, 2018 *Accepted: December 28, 2018*

Introduction

Urine dipstick analysis is one of the most commonly performed tests in clinical laboratories. It is a simple and rapid test suitable for emergency as well as for primary care settings where urine dipstick analysis is often used to diagnose urinary tract infections, proteinuria, haematuria, and some other conditions (1,2).

Unfortunately, urine dipstick testing suffers from a substantial variability among manufacturers respective to their sensitivity, specificity and measurement range (3). It has been demonstrated that some urine dipsticks have poor ability to accurately detect proteinuria due to their low sensitivity (4). Various dipsticks may differ in their diagnostic

performance regarding leukocyte and erythrocyte detection (5). There is also evidence that urine dipstick pH analysis shows insufficient accuracy (6).

Such difference between manufacturers increases the possibility for diagnostic errors, leading to inappropriate decisions thus creating a serious risk for the patient. Obviously, it is highly desirable that results of urine dipstick testing are comparable between different test strip manufacturers.

There are 195 medical laboratories in Croatia, out of which majority (N = 174) perform urine dipstick testing. Based on the data of our national External Quality Assessment (EQA) provider (Croatian Centre for Quality Assessment in Laboratory Medicine, CROQALM), there are 14 urine dipstick manufacturers on the market, who all together offer 24 different types of urine dipsticks (EQA – CROQALM laboratory reports, unpublished data). Our hypothesis was that dipsticks used for qualitative urinalysis in Croatia are heterogeneous and poorly standardized. Although many authors have studied the comparability of several dipsticks, such a comprehensive analysis of 12 different dipstick manufacturers so far has not been done. Our aim was therefore: a) to determine the level of agreement between 12 most commonly used dipsticks in Croatia using urine samples, and b) to examine their analytical performance by determining their repeatability and analytical accuracy for glucose and total protein (by comparison with quantitative measurement on chemistry analyser).

Materials and methods

Samples

This analytical validation study was done in the University Hospital "Sveti Duh" (Zagreb, Croatia) between March and May 2017. We have collected 75 urine samples from in- and out- patients to validate comparability and accuracy of 12 dipstick brands used in Croatia. Samples were collected randomly (at any time) in polystyrene tubes (10 mL, 16x95, Deltalab, Barcelona, Spain) and analysed within 2 hours of sample receipt. Additionally, 12 urine samples were used to validate repeatability for each dipstick brand. The list of 12 dipsticks used in this study is provided in Table 1.

Table 1. Most common urine dipstick brands and manufacturers in Croatia, used in this study

Number	Dipstick	Manufacturer (City, State)
1	Combur 10 Test M	Roche (Mannheim, Germany)
2	ChoiceLine 10	Roche (Mannheim, Germany)
3	Combur 10 Test UX	Roche (Mannheim, Germany)
4	ComboStik 10M	DFI Co., Ltd. (Gimhae, South Korea)
5	ComboStik 11M	DFI Co., Ltd. (Gimhae, South Korea)
6	CombiScreen 11SYS	Analyticon (Lichtenfels, Germany)
7	CombiScreen 10SL	Analyticon (Lichtenfels, Germany)
8	Combina 13	Human (Wiesbaden, Germany)
9	Combina 11S	Human (Wiesbaden, Germany)
10	Combina 10M	Human (Wiesbaden, Germany)
11	UriGnost 11	BioGnost Ltd. (Zagreb, Croatia)
12	Multistix 10SG	Siemens (Erlangen, Germany)

Urine samples were carefully chosen according to the results (negative, 1+, 2+ and 3+) obtained on automated urinalysis chemistry analyser (iChem Velocity, Beckman Coulter, Brea, USA) to ensure a wide range of concentrations of each dipstick parameter. Only urine samples with adequate volume (at least 5 mL) have been selected and further divided into three aliquots (1 mL each) and the rest of the sample was used for urine test strips dipping. Aliquotes were measured on three automated analysers to assess dipsticks accuracy for glucose and total protein. Patient data privacy was ensured throughout the study. Study was done with the approval of the hospital Ethical Committee.

Dipsticks comparability and repeatability

Comparability and repeatability of the dipsticks were performed according to the Clinical and Laboratory Standards Institute (CLSI) guideline EP12-A2 (7). The comparability of urine dipsticks was ex-

amined on 75 urine samples for parameters: glucose, total protein, erythrocytes, lekocytes, ketones, bilirubin, urobilinogen, nitrite and specific gravity (SG). Test strips were examinated visually by three observers at the same time, using the color scale provided by the manufacturer. In case when there was a disagreement between observers, a reassessment was done and final color was agreed by a consensus opinion of all three observers.

Dipsticks repeatability was tested on 20 repeated measurements of each dipstick brand. Replicates were done using the same urine sample in one laboratory (under the same ambient conditions, *e.g.* the same room temperature and light exposure). Three observers also visually examined these dipsticks.

Analytical accuracy: comparison of dipstick and quantitative measurement

Analytical accuracy assessment was performed according to CLSI EP09-A3 guideline (8). Accuracy of urine dipsticks for glucose and total protein was investigated on 75 urine samples. Glucose and total protein were quantitatively measured using three different analysers on three locations in Zagreb: AU400 (Beckman Coulter, Brea, USA) in University Hospital "Sveti Duh", Architect plus c4000 (Abbott, Abbott Park, USA) in Children's Hospital Zagreb, and Cobas 6000 c501 (Roche Diagnostics GmbH, Mannheim, Germany) in University Hospital Centre Zagreb. Urine aliquots (1 mL) were wrapped in aluminum, transported to other two laboratories on the same day and analysed within 4 hours. Urine proteins were measured with original reagents, by photometric dye-binding pyrogallol red molybdate assay on AU400 analyser, and turbidimetric method with benzethonium chloride on Cobas 600 c501 and Architect plus c4000. Glucose was measured by hexokinase method on all three analysers, with original reagents. Systems were monitored daily using commercial internal quality control (IQC) materials: AU400 (Liquichek urine chemistry control, Bio-Rad Laboratories Inc., Hercules, USA, LOT: 66781 and 66782), Architect plus c4000 (Multichem U, Technopath, New York, USA, LOT: 23110161 and 23109162) and for Cobas 600 c501 (Liquichek urine chemistry control, Bio-Rad Laboratories Inc., Hercules, USA, LOT: 66771 and 66752). Analysers were calibrated in case IQC results were out of range.

Since there is no recommendation for a reference method for urinary total protein measurement, and given the large differences between these two methods, dipstick results for proteins were compared with quantitative measurements by two methods (pyrogallol red molybdate and benzethonium chloride) separately (9). Furthermore, dipstick results for glucose were compared to mean value of all three chemistry analysers.

Day-to-day precision of glucose and total protein in urine samples

For each analyser included in this study, day-to-day precision was evaluated on measurements of two level control materials (Liquichek urine chemistry control, Bio-Rad Laboratories Inc. and Multichem U, Technopath) in 20 days. Day to day precision performance criteria (coefficient of variation: CV, %) were set in accordance with Reference Institute for Bioanalytics (RfB): for proteins 19.73% and 10.13% (at concentrations 0.15 and 0.97 g/L) and for glucose 10.94% and 7.81% (at concentrations 1.2 and 11 mmol/L).

Statistical analysis

Level of agreement between each dipstick and the reference dipstick was tested by weighted kappa test and expressed as Cohen kappa value (κ). The most commonly used brand in Croatia in 2017 (based on the data from our national EQA provider), served as a reference. Kappa value was considered acceptable if ≥ 0.80 (10). Although the number of fields for each parameter differed between the dipstick brands, for the purpose of the assessment of the agreement, the observers have merged some categories (where the number of observations was low) and results were classified into 4 categories (neg/norm (N), 1+, 2+, 3+). For each category at least 10 samples were used.

We have excluded from comparability analysis those dipstick brands which did not have concentrations assigned to categories: ChoiceLine 10 (Roche), Combur 10 Test UX (Roche), ComboStik 10M (DFI Co., Ltd.), ComboStik 11M (DFI Co., Ltd.), Combina 10M (Human) and Multistix 10SG (Siemens) for bilirubin and UriGnost 11 (BioGnost Ltd.) for erythrocytes.

Analytical accuracy of urine dipsticks for glucose and total protein was assessed by comparing the readings from the dipsticks with the true value of the parameter measured by the quantitative test results from chemistry analysers. Glucose and total protein concentrations were distributed into categories: for total protein: N = 0 - 0.29 g/L, 1 = 0.30 - 0.99 g/L, 2 = 1.00 - 2.99 g/L, 3 = more than 3.00 g/L); and for glucose: N = 0 - 2.79 mmol/L, 1 = 2.80 - 8.29 mmol/L, 2 = 8.30 - 27.99 mmol/L, 3 = more than 28 mmol/L. Categories obtained by dipstick and quantitative testing were compared and number of true positive and negative, and false positive and negative findings were established. According to these results, analytical sensitivity and specificity were calculated for each dipstick brand. Dipsticks with sensitivity and specificity ≥ 90% were considered excellent, those with ≥ 80% were satisfactory and the other dipsticks (< 80%) were considered as being of less than acceptable quality. Acceptance criteria for repeatability was 90% (18/20 results) of repeated measurements.

Data were analysed using MedCalc 12.6.2.0 (Ostend, Belgium) statistical software.

Results

Dipsticks comparability

Combur 10 Test M (Roche) was chosen as a reference because it was the most commonly used dipstick brand in Croatia in 2017 according to the national EQA provider (44/174, 25%). Levels of agreement between dipsticks and the reference for each parameter, expressed as κ, are shown in Table 2. Combur 10 Test UX (Roche) showed the best agreement with the reference dipstick (κ > 0.80) for all parameters. The lowest level of agreement was shown for Combina 13 (Human) and the reference, particularly for bilirubin, urobilinogen, pH and SG (κ < 0.46).

The best overall comparability (κ > 0.80) was achieved for glucose and nitrite (11/11 brands) and total protein (10/11 brands). Moderate agreement

Table 2. Agreement between 11 most common dipstick brands in Croatia with the reference Combur 10 Test M (Roche)

	kappa value									
Dipstick	**Glc**	**Prot**	**Erc**	**Leu**	**Ket**	**Bil**	**Ubg**	**Nit**	**pH**	**SG**
ChoiceLine 10	0.90	0.89	0.76	0.82	0.73	/	0.89	0.97	0.71	0.81
Combur 10 Test UX	0.99	0.93	0.94	0.85	0.92	/	0.90	0.97	0.95	0.90
ComboStik 10M	0.89	0.87	0.75	0.71	0.71	/	0.51	0.97	0.40	0.31
ComboStik 11M	0.86	0.87	0.72	0.78	0.69	/	0.46	0.97	0.43	0.32
CombiScreen 11SYS	0.90	0.87	0.79	0.71	0.71	0.54	0.78	1.00	0.87	0.64
CombiScreen 10SL	0.89	0.87	0.76	0.70	0.80	0.51	0.74	1.00	0.87	0.62
Combina 13	0.84	0.79	0.60	0.71	0.84	0.16	0.36	0.97	0.46	0.42
Combina 11S	0.88	0.81	0.76	0.68	0.71	0.44	0.81	0.97	0.79	0.60
Combina 10M	0.91	0.87	0.78	0.72	0.80	/	0.19	1.00	0.53	0.41
UriGnost 11	0.83	0.87	/	0.78	0.85	0.33	0.85	0.97	0.88	0.49
Multistix 10SG	0.80	0.87	0.71	0.71	0.78	/	0.89	0.97	0.56	0.54

Darker grey fields represent the highest κ-values (κ ≥ 0.80); lighter grey fields show lower κ-values (κ < 0.80); white fields represent excluded parameters (/). Glc – glucose. Prot – total protein. Erc – erythrocytes. Leu – lekocytes. Ket – ketones. Bil – bilirubin. Ubg – urobilinogen. Nit – nitrite. pH - acidity or basicity. SG – specific gravity.

(κ = 0.60 - 0.79) was observed for erythrocytes (9/10 brands) and leukocytes (9/11 brands). Overall, lowest kappa values were achieved for bilirubin. There was a weak level of agreement (κ = 0.44 - 0.54) for bilirubin in 3/5 brands and for the other two brands the agreement was minimal to none (κ = 0.33 - 0.16).

Dipsticks repeatability

Repeatability was assessed on 20 replicates of each dipstick brand (Table 3). Repeatability for at least one parameter was < 90% for 6/12 dipstick brands. The most problematic parameter was pH, where as many as three dipstick brands had < 90% repeatability: ChoiceLine 10 (Roche), CombiScreen 10SL (Analyticon) and Combina 13 (Human).

Day-to-day precision of glucose and total protein in urine samples

Day-to-day precision (CV, %) for total protein measurement ranged 1.90 – 3.90% in the lower range (concentrations 0.18 – 0.27 g/L) and 1.10–2.88% in the higher range concentrations (0.62 – 1.26 g/L) on all three analysers. For urinary glucose measurement, CVs were 1.60 – 3.29% at lower concentrations (1.43 – 1.89 mmol/L) and 1.21 – 1.71% at higher concentrations (16.28 – 20.40 mmol/L) of control materials on all three analysers.

Analytical accuracy: comparison of dipstick and quantitative measurement

Glucose

Analytical sensitivity and specificity of each dipstick for urinary glucose measurement is presented in Table 4. While sensitivity for glucose was > 90% for 5/12 dipstick brands, their specificity was modest (71 - 83%). Only three dipstick brands, Combina 13 (Human), Urignost 11 (BioGnost Ltd.) and Multistix 10SG (Siemens), were able to detect glucose with high specificity (> 90%), but with much lower sensitivity and higher false negative rate.

Proteins

Analytical accuracy for urinary proteins is presented for each method (pyrogallol red and benzethonium chloride) separately (Table 5). Regarding pyrogallol red molybdate assay (AU 400, Beckman Coulter), none out of twelve dipsticks detected proteins with analytical sensitivity or specificity > 80%. Sensitivity was the highest (75%) for Combi-

Table 3. Repeatability of 12 most common dipstick brands in Croatia (assessed on 20 replicates for all parameters).

	Number of acceptable replicates / total number of replicates									
Dipstick	**SG**	**pH**	**Leu**	**Nit**	**Prot**	**Glc**	**Ket**	**Bil**	**Ubg**	**Erc**
Combur 10 Test M	20/20	19/20	20/20	20/20	20/20	20/20	20/20	20/20	20/20	20/20
ChoiceLine 10	19/20	17/20	20/20	20/20	20/20	20/20	20/20	20/20	20/20	20/20
Combur 10 Test UX	20/20	20/20	20/20	20/20	20/20	20/20	20/20	20/20	20/20	20/20
ComboStik 10M	20/20	20/20	19/20	20/20	20/20	20/20	20/20	20/20	20/20	20/20
ComboStik 11M	20/20	20/20	19/20	20/20	20/20	18/20	18/20	20/20	20/20	20/20
CombiScreen 11SYS	16/20	20/20	20/20	20/20	19/20	18/20	15/20	20/20	20/20	20/20
CombiScreen 10SL	20/20	16/20	20/20	20/20	20/20	20/20	20/20	20/20	20/20	20/20
Combina 13	19/20	11/20	20/20	20/20	19/20	20/20	20/20	19/20	20/20	18/20
Combina 11S	18/20	20/20	20/20	20/20	19/20	20/20	20/20	20/20	20/20	20/20
Combina 10M	20/20	20/20	18/20	20/20	19/20	20/20	20/20	20/20	20/20	20/20
UriGnost 11	19/20	20/20	17/20	20/20	20/20	20/20	20/20	18/20	20/20	20/20
Multistix 10SG	20/20	18/20	20/20	20/20	20/20	20/20	20/20	13/20	20/20	20/20

Grey fields represent parameters that did not meet the acceptance criteria. SG – specific gravity. pH – acidity or basicity. Leu – lekocytes. Nit – nitrite. Prot – proteins. Glc – glucose. Ket – ketones. Bil – bilirubin. Ubg – urobilinogen. Erc – erythrocytes.

Table 4. The analytical sensitivities and specificities for glucose for 12 most common dipsticks in Croatia with hexokinase method as a reference

Dipstick	Manufacturer	Sensitivity	Specificity
Combur 10 Test M	Roche	97.0%	81.0%
ChoiceLine 10	Roche	96.3%	75.0%
Combur 10 Test UX	Roche	97.0%	83.3%
ComboStik 10M	DFI Co., Ltd.	80.0%	80.0%
ComboStik 11M	DFI Co., Ltd.	73.3%	80.0%
CombiScreen 11SYS	Analyticon	89.3%	76.6%
CombiScreen 10SL	Analyticon	85.7%	76.6%
Combina 13	Human	69.7%	92.9%
Combina 11S	Human	95.8%	70.6%
Combina 10M	Human	93.1%	80.4%
UriGnost 11	BioGnost Ltd.	72.7%	97.6%
Multistix 10SG	Siemens	67.7%	93.2%

Grey fields represent acceptable sensitivity or specificity (light grey fields ≥ 80%, darker grey > 90%).

Table 5. The analytical sensitivities and specificities for urinary total protein for 12 most common dipsticks in Croatia with pyrogallol red molybdate assay and turbidimetric method with benzethonium chloride as a references

		Pyrogallol red molybdate assay		Turbidimetric method with benzethonium chloride	
Dipstick	Manufacturer	Sensitivity	Specificity	Sensitivity	Specificity
Combur 10 Test M	Roche	69.8%	75.0%	87.2%	72.2%
ChoiceLine 10	Roche	69.2%	69.4%	81.8%	64.3%
Combur 10 Test UX	Roche	66.7%	66.7%	85.7%	65.0%
ComboStik 10M	DFI Co., Ltd.	60.0%	71.4%	77.8%	69.2%
ComboStik 11M	DFI Co., Ltd.	60.0%	71.4%	77.8%	69.2%
CombiScreen 11SYS	Analyticon	61.5%	69.4%	75.8%	66.7%
CombiScreen 10SL	Analyticon	60.0%	71.4%	73.5%	68.3%
Combina 13	Human	41.0%	72.2%	55.9%	70.7%
Combina 11S	Human	75.0%	45.7%	85.3%	41.5%
Combina 10M	Human	70.0%	62.9%	91.7%	66.7%
UriGnost 11	BioGnost Ltd.	70.7%	70.6%	86.1%	66.7%
Multistix 10SG	Siemens	67.5%	74.3%	80.0%	67.5%

Light grey fields represent the highest (≥ 80%) and dark grey fields the lowest (< 60%) sensitivities and specificities.

na 11S (Human), but this dipstick brand had lowest specificity (only 45%). Specificity was the highest (75%) for Combur 10 Test M (Roche), but its sensitivity was average (70%). Combina 13 (Human) had the lowest sensitivity for proteins (41%) and the highest false negative rate. Ability of other dipsticks to detect proteins specifically, varied between 63 - 74%.

As of the analytical accuracy respective to the turbidimetric method with benzethonium chloride, Combina 10M (Human) had the highest analytical sensitivity (92%) and several other dipsticks have achieved sensitivity > 80%. However, analytical specificities for these dipsticks varied between 41 – 72%. Combina 11S (Human) had the lowest specificity for proteins (42%) and the highest false positive rate (24/75). The lowest sensitivity (56%) was observed for Combina 13 (Human), with the highest false negative rate (15/75) and only average specificity (71%).

Discussion

In this study, we performed comprehensive analytical verification of 12 most commonly used dipsticks in Croatia. Our results showed that these dipsticks are not sufficiently comparable and that they vary in analytical performance. Agreement between the dipsticks was acceptable for nitrites, proteins and glucose but there was remarkable diversity for other parameters like bilirubin, urobilinogen, pH and specific gravity. The most important clinically relevant finding was that most of the dipsticks did not accurately detected glucose and proteins.

As previously described in the literature, quantitative methods for urinary proteins are not mutually comparable and none of the available methods is considered as a "gold standard" method (9). In our study, the agreement of dipsticks was better with turbidimetric method for total urinary protein. Respective to pyrogallol red molybdate assay, none of the dipsticks showed acceptable accuracy for total urinary protein. On the other hand, respective to turbidimetric method with benzethonium chloride, seven out of twelve dipsticks showed satisfactory sensitivity but were lacking the adequate specificity for urinary proteins. Consistent with these observations, reference intervals for total urinary protein excretion recommended by the European Urinalysis Group are higher for pyrogallol red molybdate assay (< 180 mg/day) than turbidimetric methods (< 75 mg/day) (11).

In general, our results demonstrate that dipsticks have unacceptably high false negative rates and even higher false positive rates for total protein. Our findings are in line with several previous studies, who have also confirmed the suboptimal accuracy of qualitative urine dipstick analysis for total urinary protein (4,12). Our findings also point to low accuracy of urine dipstick analysis for glucose. Only four dipstick brands have achieved both sensitivity and specificity higher than 80%. This is in line with some earlier observations (13). Considering this limitation, International Diabetes Federation suggests the use of glucose dipstick testing only in low resource settings, where other glucose tests are not affordable (14). Obviously, substantial improvement of the accuracy of dipsticks for protein and glucose is highly warranted.

Whereas the level of agreement between the dipsticks in our study was acceptable for nitrites, it was less than acceptable for erythrocytes and leukocytes. Given the widespread heterogeneity of available brands of dipstick manufacturers in Croatia, and probably even worldwide, such lack of agreement between various manufacturers creates the opportunity for patient misclassification in these conditions where parameters such as nitrites, erythrocytes and leukocytes are of diagnostic relevance (*e.g.* urinary tract infections). Moreover, at least for some manufacturers, low reproducibility for leukocytes might be an additional issue. Urine dipstick testing (especially the combination of leukocytes, blood and nitrites) has been proposed as a first step to diagnose urinary tract infection (UTI) (15,16). National Institute for Health and Care Excellence (NICE) guidelines recommend using dipsticks as a screening tool, based on the assumption that UTI can be safely ruled out with both negative leukocyte esterase and nitrite in asymptomatic patients (17). Obviously, while this may be the case for some dipsticks, other may not be as accurate. Therefore, unless some improvement in this respect is made, it is to be expected that at least for the users of some dipstick manufacturers, the ability to detect UTI will remain less that acceptable. This is even more worrying, given the fact that positive leukocytes in extravascular fluids such as ascites and synovial fluid have re-

cently been proposed as useful indication for some conditions like spontaneous bacterial peritonitis and periprosthetic joint infection, respectively (18-22).

Low level of agreement of urine dipstick parameters is an issue in some other health conditions where erythrocytes alone are used in diagnostic process. For example, dipstick blood assessment is often used for bladder cancer regular check-up. NICE guidelines state that asymptomatic microhaematuria may be an early sign of a bladder cancer in people aged 60 and older, but do not define whether dipsticks or microscopy should be used for asymptomatic microhaematuria assessment (23). Moreover, American Urological Association recommends that positive blood on the dipstick and negative on sediment count, should be followed by three additional sediment microscopic evaluations. If at least one of those tests is positive, further actions and treatment decisions should be taken (24). Apparently, the above-mentioned guidelines and recommendations do not take into account the low accuracy of dipstick testing for erythrocytes (haematuria) and low level of agreement between various manufacturers, and thus may lead to either over- or under-estimation of the occurrence of haematuria, which may significantly jeopardize patient safety. Due to unacceptable high false negative rate, negative dipstick test cannot rule out disease of symptomatic patients. False positive haematuria dipstick result can also lead to increased number of microscopic sediment examinations, further urological examinations and unnecessary testing like imaging or cystoscopy (25). Hence, high false positive rate of erythrocytes may also substantially increase laboratory workload and affect healthcare costs. Given the reasons discussed above, it is essential that dipstick manufacturers improve analytical performance for dipstick ability to accurately detect erythrocytes in urine. Otherwise, it is reasonable to consider diagnostic value of blood on the dipstick quite limited or even questionable.

In our study on 12 most common dipsticks in Croatia there was a wide heterogeneity in kappa values for bilirubin, urobilinogen, pH and specific gravity, pointing to the low comparability of the results obtained by different brands of dipsticks. Also, some dipsticks in our study were of unacceptable repeatability for pH. Some previous literature reports have also demonstrated unacceptable precision and accuracy of the dipsticks comparing them with gold standard, pH – meter (26). It has also been reported that dipsticks vary in accuracy due to proportions and combinations of the reagents (like methyl red and bromthymol blue) in pH fields provided by different manufacturers (27). Previous studies described usefulness of specific gravity as additional parameter which increases the accuracy for proteinuria assuming that concentrated urine is more likely to have positive protein field on the dipstick (28). Hillege opposed this statement claiming that this algorithm has nonsignificant yield in diagnostic accuracy (29). Furthermore, there is inconsistency in some earlier studies which described the use of specific gravity in evaluating the degree of dehydration and optimal urine output in patients with nephrolithiasis (30). Although bilirubin and urobilinogen in urine indicate several liver conditions like hepatocellular disease, biliary obstruction and cholestatic jaundice, it should be noted that liver diseases are diagnosed after clinical examination, some obvious symptoms like yellow skin and eye discoloration, imaging studies and liver tests in blood. Therefore, bilirubin and urobilinogen dipstick tests have no real diagnostic value (11). Given the low analytical quality and limited clinical utility of these parameters, it would be reasonable to question the need for these parameters in the first place.

Our study has some potential limitations. We have assessed the level of agreement of 12 most common dipstick brands by comparing them to the one which was the most common in Croatia. It could be that the agreement would be different if some other manufacturer was chosen as a reference. Also, we have analyzed dipstick repeatability by testing different urine sample for every dipstick brand, since it was logistically challenging to ensure an adequate amount of urine to do all testing in the same urine. We acknowledge this as a limitation and potential source of bias, due to matrix effects. Furthermore, only pathological samples were chosen for this part of the study thus possi-

ble endogenous and exogenous interferences could have also affected our results. Finally, we have assessed the accuracy only for glucose and proteins. We acknowledge that it would be beneficial to also evaluate the accuracy for some other parameters, such as leukocytes, erythrocytes and nitrites, by comparison with urine sediment microscopy and microbiological testing. Nevertheless, due to some local challenges and operational difficulties we were not able to perform such analysis in this study.

In summary, 12 most commonly used dipsticks in Croatia showed low level of agreement among each other. Dipsticks accuracy and precision showed considerable variability between different manufacturers. Most dipsticks do not accurately detect glucose and proteins. Given the widespread heterogeneity of available brands of dipstick manufacturers in Croatia, but also possibly even worldwide, these issues create the opportunity for patient misclassification, jeopardize patient safety and increase healthcare costs. Obviously, some improvement in that respect (*i.e.* standardization among manufacturers and improvement of the quality of dipsticks) is highly necessary to minimize patient risk. We believe that, although our study addresses the situation in Croatia, it is also relevant to other countries in Europe and beyond.

Potential conflict of interest

None declared.

References

1. *Stein R, Dogan HS, Hoebeke P, Kočvara R, Nijman RJ, Radmayr C, et al. European Association of Urology; European Society for Pediatric Urology. Urinary tract infections in children: EAU/ESPU guidelines. Eur Urol. 2015;67:546-58. https://doi.org/10.1016/j.eururo.2014.11.007*
2. *Matulewicz RS, DeLancey JO, Pavey E, Schaeffer EM, Popescu O, Meeks JJ. Dipstick Urinalysis as a Test for Microhaematuriaand Occult Bladder Cancer. Bladder Cancer. 2017;3:45-9. https://doi.org/10.3233/BLC-160068*
3. *Correa ME, Côté AM, De Silva DA, Wang L, Packianathan P, von Dadelszen P, et al. Visual or automated dipstick testing for proteinuria in pregnancy? Pregnancy Hypertens. 2017;7:50-3. https://doi.org/10.1016/j.preghy.2017.01.005*
4. *Kumar A, Kapoor S, Gupta RC. Comparison of urinary protein: Creatinine index and dipsticks for detection of microproteinuria in diabetes mellitus patients. J Clin Diagnostic Res. 2013;7:622–6. https://doi.org/10.7860/JCDR/2013/4745.2867*
5. *Ko K, Kwon MJ, Ryu S, Woo HY, Park H. Performance Evaluation of Three URiSCAN Devices for Routine Urinalysis. J Clin Lab Anal. 2016;30:424-30. https://doi.org/10.1002/jcla.21874*
6. *Abbott JE, Miller DL, Shi W, Wenzler D, Elkhoury FF, Patel ND, et al. Optimization of urinary dipstick pH: Are multiple dipstick pH readings reliably comparable to commercial 24-hour urinary pH? Investig Clin Urol. 2017;58:378-82. https://doi.org/10.4111/icu.2017.58.5.378*
7. *Clinical and laboratory standards institute (CLSI). User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline - Second Edition. CLSI Document EP12-A2. Wayne, PA:CLSI,2008.*
8. *Clinical and laboratory standards institute (CLSI). Measurement Procedure Comparison and Bias Estimation Using Patient Samples - Third Edition. CLSI Document EP09-A3. Wayne, PA:CLSI,2013.*
9. *Yalamati P, Karra ML, Bhongir AV. Comparison of Urinary Total Proteins by Four Different methods. Ind J Clin Biochem. 2016;31:463-7. https://doi.org/10.1007/s12291-016-0551-3*
10. *McHugh ML. Interrater reliability: the kappa statistic, Biochem Med (Zagreb). 2012;22:276–82. https://doi.org/10.11613/BM.2012.031*
11. *Kouri TT, Gant VA, Fogazzi GB, Hofmann W, Hallander HO, Guder WG. Towards European urinalysis guidelines. Clin Chim Acta. 2000;297:305–11. https://doi.org/10.1016/S0009-8981(00)00256-4*
12. *White SL, Yu R, Craig JC, Polkinghorne KR, Atkins RC, Chadban SJ. Diagnostic accuracy of urine dipsticks for detection of albuminuria in the general community. Am J Kidney Dis. 2011;58:19-28. https://doi.org/10.1053/j.ajkd.2010.12.026*
13. *Storey HL, van Pelt MH, Bun S, Daily F, Neogi T, Thompson M, et al. Diagnostic accuracy of self-administered urine glucose test strips as a diabetes screening tool in a low-resource setting in Cambodia. BMJ Open. 2018;8:e019924. https://doi.org/10.1136/bmjopen-2017-019924*
14. *International Diabetes Federation. Clinical Guidelines Task Force Global Guideline for Type 2 Diabetes. Available at: https://www.idf.org/e-library/guidelines/79-global-guideline-for-type-2-diabetes. Accessed April 5th 2018.*
15. *Grabe M, Johansen BTE, Botto H, Çek M, Naber KG, Pickard RS, et al. Guidelines on Urological Infections. Available at: https://uroweb.org/wp-content/uploads/19-Urological-infections_LR2.pdf. Accessed April 5th 2018.*

16. Fernandes DJ, Jaidev MD, Castelino DN. Utility of dipstick test (nitrite and leukocyte esterase) and microscopic analysis of urine when compared to culture in the diagnosis of urinary tract infection in children. Int J Contemp Pediatr. 2018;5:156-60. https://doi.org/10.18203/2349-3291.ijcp20175578
17. National Institute for Health and Care Excellence (NICE). Urinary tract infection in under 16s Diagnosis and management; Clinical guideline. Available at: https://www.nice.org.uk/guidance/cg54. Accessed April 6th 2018.
18. Rathore V, Joshi H, Kimmatkar DP, Malhotra V, Agarwal D, Beniwal P, et al. Leukocyte Esterase Reagent Strip as a Bedside Tool to Detect Peritonitis in Patients Undergoing Acute Peritoneal Dialysis. Saudi J Kidney Dis Transpl. 2017;28:1264-9. https://doi.org/10.4103/1319-2442.220875
19. Chugh K, Agrawal Y, Goyal V, Khatri V, Kumar P. Diagnosing bacterial peritonitis made easy by use of leukocyte esterase dipsticks. Int J Crit Illn Inj Sci. 2015;5:32-7. https://doi.org/10.4103/2229-5151.152337
20. Oey RC,Kuiper JJ, van Buuren HR, de Man RA. Reagent strips are efficient to rule out spontaneous bacterial peritonitis in cirrhotics. Neth J Med. 2016;74:257-61.
21. Wang C, Li R, Wang Q, Duan J, Wang C. Leukocyte Esterase as a Biomarker in the Diagnosis of Periprosthetic Joint Infection. Med Sci Monit. 2017;23:353-8. https://doi.org/10.12659/MSM.899368
22. Tischler HE, Cavanaugh KP, Parvizi J. Leukocyte Esterase Strip Test: Matched for Musculoskeletal Infection Society Criteria. J Bone Joint Surg Am. 2014;96:1917-20. https://doi.org/10.2106/JBJS.M.01591
23. National Institute for Health and Care Excellence (NICE). Suspected cancer: recognition and referral; NICE guidelines. Available at: https://www.nice.org.uk/guidance/ng12. Accessed April 6th 2018.
24. Davis R, Jones JS, Barocas DA, Castle EP, Lang EK, Leveillee RJ, et al. Diagnosis, evaluation and follow-up of asymptomatic microhaematuria(AMH) in adults: AUA guideline. J Urol. 2012;188:2473-81. https://doi.org/10.1016/j.juro.2012.09.078
25. Linder BJ, Bass EJ, Mostafid H, Boorjian SA. Guideline of guidelines: asymptomatic microscopic haematuria. BJU International. 2018;121:176–83. https://doi.org/10.1111/bju.14016
26. Ilyas R, Chow K, Young JG. What Is the Best Method to Evaluate Urine pH? A Trial of Three Urinary pH Measurement Methods in a Stone Clinic. J Endourol. 2015;29:70-4. https://doi.org/10.1089/end.2014.0317
27. Desai RA, Assimos DG. Accuracy of Urinary Dipstick Testing for pH Manipulation Therapy. J Endourol. 2008;22:1367-70. https://doi.org/10.1089/end.2008.0053
28. Constantiner M, Sehgal AR, Humbert L, Constantiner D, Arce L, Sedor JR, et al. A dipstick protein and specific gravity algorithm accurately predicts pathological proteinuria. Am J Kidney Dis. 2005;45:833-41. https://doi.org/10.1053/j.ajkd.2005.02.012
29. Hillege HL. Can an algorithm based on dipstick urine protein and urine specific gravity accurately predict proteinuria? Nat Clin Pract Nephrol. 2006;2:68-9. https://doi.org/10.1038/ncpneph0099
30. Khorami MH, Hashemi R, Bagherian-Sararoudi R, Sichani MM, Tadayon F, Shahdoost AA, et al. The assessment of 24-h urine volume by measurement of urine specific gravity with dipstick in adults with nephrolithiasis. Adv Biomed Res. 2012;1:86. https://doi.org/10.4103/2277-9175.105168